Vol. 38, No. 4 August, 1948

Science Library

Journal of Experimental Psychology

AUG 26 1948

EDITED BY

FRANCIS W. IRWIN

UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CLARENCE H. GRAHAM, *Columbia University*
HARRY HELSON, *Bryn Mawr College*
DONALD B. LINDSLEY, *Northwestern University*
DONALD G. MARQUIS, *University of Michigan*
GEORGE A. MILLER, *Harvard University*
L. L. THURSTONE, *University of Chicago*
G. R. WENDT, *University of Rochester*
JOSEPH ZUBIN, *N. Y. Psychiatric Institute, Columbia University*

CONTENTS

The Effect of Low Intensities of Hunger on the Behavior Mediated by a Habit of Maximum Strength: I. SALTZMAN AND S. KOCH 347

The Effect of Success and Failure upon the Attractiveness of Activities as a Function of Experience, Expectation, and Need: MILDRED E. GEBHARD 371

Stimulus Identity as Related to Response Specificity and Response Generalization: D. D. WICKENS 389

Measurement of Accommodation and Convergence Time as Part of a Complex Visual Adjustment: R. C. TRAVIS 395

A Behavioral Analysis of Degree of Reinforcement and Ease of Shifting to New Responses in a Weigl-Type Card-Sorting Problem: D. A. GRANT AND ESTA A. BERG 404

Spatial Orientation in the White Rat: H. C. WILCOXON AND R. H. WATERS 412

The Influence of Intensity of Unconditioned Stimulus upon Acquisition of a Conditioned Response: G. E. PASSEY 420

'Spontaneous Recovery' of Verbal Associations: B. J. UNDERWOOD 429

Anxiety-Produced Interference in Serial Rote Learning with Observations on Rote Learning after Partial Frontal Lobectomy: R. B. MALMO AND A. AMSEL 440

Studies in Space Orientation. II. Perception of the Upright with Displaced Visual Fields and with Body Tilted: S. E. ASCH AND H. A. WITKIN 455

Pitch Characteristics of Short Tones. II. Pitch as a Function of Tonal Duration: J. M. DOUGHTY AND W. R. GARNER 478

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STS., LANCASTER, PA.

AND 1515 MASSACHUSETTS AVE., N. W., WASHINGTON 5, D. C.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879
Acceptance for mailing at the special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, Section 538, P. L. & R., authorized September 11, 1947.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published bi-monthly by the American Psychological Association, Inc. The annual subscription rate is $7.00 (foreign, $7.25), single copies, $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave., N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Articles of a length exceeding 25 printed pages cannot be accepted.

With rare exceptions, articles are published in the order of their receipt. Authors are supplied with 50 free offprints, without covers, in the case of articles of seven or more printed pages, and with five free copies of the issue in the case of shorter contributions. One-half of the cost of tables, cuts, formulas, and author's alterations in galley proof is charged to the author.

Immediate publication can be given to articles whose authors assume the full cost of publication.

Address articles submitted for publication to the editor, Francis W. Irwin, Department of Psychology, University of Pennsylvania, Philadelphia 4, Pa.

This JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. ANDERSON and W. L. VALENTINE, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, **41**, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *E*s, *E*'s, and *E*s'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts should be typed double-spaced throughout, *including tables, footnotes, and bibliography*. Tables, figures, and footnotes should be placed on separate pages at the end of the manuscript.

Journal of
Experimental Psychology

VOL. 38, NO. 4 AUGUST, 1948

THE EFFECT OF LOW INTENSITIES OF HUNGER ON THE BEHAVIOR MEDIATED BY A HABIT OF MAXIMUM STRENGTH [1]

BY IRVING SALTZMAN AND SIGMUND KOCH

Indiana University and Duke University

INTRODUCTION

The present experiment seeks to determine the relation between low intensities of hunger, ranging from one-half to two hours, and the vigor of performance of a simple habit, already learned to the limit of practice. So far as the writers know, no previous quantitatively oriented experiment has measured the effect on behavior strength of drive states comparably weak in intensity to those used in the present study. In consequence, this experiment should be of value in providing a first empirical estimate of the influence on behavior of drive intensities which *seem* so slight that their dynamic role is usually considered negligible, or non-existent. At the same time, this study will serve further to specify the nature of the general relationship between drive intensity and behavior strength. In this connection, the present investigation falls into a context of research established by two earlier experiments (Perin, **6**, and Koch and Daniel, **4**).

The Perin experiment constitutes the first quantitative attempt to determine the effect on behavior strength of varying intensities of a present drive, as combined with varying strengths of a previously learned habit. The main theoretical purpose of this experiment was to define the multiple functional relationship linking together three of the principal constructs in Hull's theory of behavior (**2**): 'habit strength' and 'drive intensity' on the one

[1] This article is based on experimental work incorporated in a thesis submitted by Dr. Saltzman to Duke University, in partial fulfillment of the requirements for the degree of Master of Arts, 1945. The second author served as thesis-director, and assumed primary responsibility for the preparation of the present report.

hand, and 'reaction potential' on the other.[2] In the terms of the experiment, habit strength corresponded to the number of previous reinforcements of a Skinner-box, lever-depression reaction, drive intensity was varied by means of differing periods of food privation after satiation, and reaction potential was measured in three ways: (1) number of responses preceding a five min. extinction criterion, (2) time preceding attainment of the criterion, and (3) latencies of the first three extinction responses.

Twelve groups of 40 rats each were given varying numbers of reinforcements, ranging from five to 90, of the bar-depression habit.[3] During this original training, hunger intensity was held at a constant level of 23 hours. The habit was then extinguished in all groups, under degrees of hunger ranging from one to 23 hours. The combinations of habit strength and drive intensity at the time of extinction, for the different groups, were selected in such a way as to make possible the determination of the following relationships: (1) reaction potential as a function of habit strength, when drive intensity is held constant at two different levels (three and 22 hours), and (2) reaction potential as a function of drive intensity, when the number of reinforcements is held constant (at 16). Equations were fitted to empirical curves plotted to represent these relationships. The final step was to combine these equations into a single equation expressing reaction potential as a joint function of habit strength and drive intensity.

Hull has used the results of this experiment as the empirical basis for one of the key postulates in his theory of behavior—'postulate 7' on *primary motivation* (2, p. 253).[4] The most general verbal description of Perin's equations, and thus of postulate 7, is that reaction potential is a *multiplicative* function of habit strength and drive intensity. More concretely, Perin's major findings may be paraphrased as follows: reaction potential increases as a positive growth function of habit strength, the asymptote of the growth function being, in turn, a slightly, positively accelerated function of drive intensity.[5]

[2] A terminological note may be helpful at this point. Definitions of the constructs 'habit strength,' 'drive intensity,' and 'reaction potential,' may be found in Hull's *Principles of behavior* (2). Although Perin's experiment was oriented towards defining the interrelations among these three variables, his report was formulated in terms of the concrete laboratory operations used to realize these variables in the experiment. Thus, Perin's study is entitled: 'Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction.' As Hull has made clear (2, Ch. XIV), 'behavior potentiality' corresponds to 'reaction potential,' 'amount of training' to 'habit strength,' and 'degree of hunger' to 'drive intensity.' In the text, we take the liberty of summarizing the Perin experiment directly in terms of its theoretical status within Hull's system, rather than in the theoretically neutral terms of Perin's original report. In the formulation of both the Koch-Daniel experiment, and the present one, the expressions, 'behavior strength' and 'vigor of behavior,' have been introduced as rough equivalents for 'reaction potential.' This has been done in order to avoid any implication that the results of these experiments are relevant exclusively, or even primarily, to Hull's theory. Rather, we wish to stress that our results embody empirical relationships relevant to any theory of behavior.

[3] Actually, Perin ran only eight of these groups. Data for the other four groups were provided by an earlier experiment conducted by Williams (9).

[4] It would appear that the Perin results serve also as the main empirical basis for a number of other assumptions involved in Hull's analysis of 'primary motivation' (cf. footnote 17).

[5] Since the three measures of reaction potential give different and non-intertranslatable values, Perin's final results are stated in the form of three separate equations. These equations

Perin's equations give rise to a rather surprising implication. Even when drive intensity is zero, reaction potential remains a positive growth function of habit strength, approximating an asymptote of substantial magnitude. In other words, even when the drive of original learning is completely absent, there should remain a definite and strong tendency toward the performance of the learned habit. The strength of this presumptive behavior tendency can be appreciated by noting that the asymptote of the reaction potential curve for *zero* hunger is approximately 28 percent of the asymptote of the curve for 22 hours of hunger (nearly the maximal level of hunger included in the experiment).

So marked a tendency toward the functioning of a learned habit, despite the absence of either the drive of original learning or any other apparent source of relevant motivation, seems incompatible with any theory which fully recognizes the dynamic role of motivation in behavior. The inherent implausibility of this finding is heightened by the fact that it was arrived at by extrapolation. Not a single group of animals was actually run under conditions of zero hunger (complete satiation).

In line with the preceding considerations, Koch and Daniel (4) made a direct empirical measurement of reaction potential at maximum habit strength and zero concurrent hunger, using a procedure patterned on Perin's in all essential respects. Twenty-six rats were allowed 70 reinforcements of the bar depression reaction, while the possibility that this amount of training did not sufficiently approximate maximum habit strength was checked by allowing six rats 90 reinforcements. Hunger was maintained at the standard intensity of 23 hours during this training. On the next day, all animals were satiated, and immediately thereafter, when the hunger drive was as close to zero as could be experimentally approached, the bar reaction was extinguished.

The Perin-Hull theoretical extrapolation demands a median value for the number of extinction responses, at zero drive and 70 reinforcements, of 16.27. The value yielded by experimental test was 1.00.[6] Perin's equation demands that the ratio of the median number of extinction responses at zero hunger and 70 reinforcements, to the median value at 22 hours' hunger and 70 reinforcements, be almost .28. Experimental measurement gives a ratio of 1.75. Qualitative observation provided grounds for believing that even this very low obtained value is inflated. The few bar reactions which occurred during the extinction tests differed markedly, in form and behavior context, from the training reactions. It thus seemed improbable that the *same* habit structures were operating during both training and extinction. If this is the case, the true reaction potential value suggested by our results is zero or below.

The flagrant inaccuracy of the Perin-Hull function at the zero drive level contains several pressing invitations to further experiment. Perhaps

all have the general form indicated by the verbal paraphrase in the text, but they differ in their specific parameters.

[6] It will be recalled that Perin used two additional measures of reaction potential: time preceding extinction, and the latencies of the first three extinction responses. The discrepancies between the median obtained values for these measures, and the corresponding Perin-Hull predictions, were even greater than in the case of the number of extinction responses.

the most obvious of these relates to the effects on reaction potential of weak intensities of hunger. The region of the Perin-Hull function between zero and three hours of hunger contains only one data plot. This represents a group extinguished under one hour of hunger and after 16 reinforcements. The gross inaccuracy of the zero drive values, together with the sparseness of data in the zero to three hour interval of the relationship, strongly suggests the desirability of checking various points within this interval by experiment. Indeed, no general formulation of the effect of hunger intensity on reaction potential, at given levels of habit strength, can legitimately be attempted without further determination of the effects of hunger intensities below three hours. As Koch and Daniel have already indicated:

> An examination of Perin's three dimensional, graphical representation of his equations . . . discloses that each point of the zero drive reinforcement curve serves as the origin of a curve of reaction potential as a function of hunger at the corresponding number of reinforcements. Each one of these hunger curves shows slight positive acceleration, originating from a positive reaction potential value (at all number of reinforcements after five), and terminating in a much higher reaction potential value at 22 hours' hunger. Our results suggest a radical revision in the form of these curves. Assuming that the reaction potential values at higher intensities of hunger are substantially correct, each hunger curve must have its origin at a reaction potential value of zero or below, rise rapidly to some point between one and three hours of hunger, and then continue, with the original slight positive acceleration, to the 22 hour hunger level. It is obviously impossible to specify the form of the initial part of the hunger curve without an empirical determination of reaction potential at a number of points at very low intensities of hunger (between, say, zero and three hours). If Perin's reaction potential values at higher strengths of hunger are correct, this revision of the hunger curve would reveal an interesting fact about the effect of very low intensities of hunger (and, possibly, drive in general) on behavior; namely, that slight increments of hunger immediately after satiation (zero hunger) have a greater relative energizing effect on behavior than similar increments taken at any higher intensity of hunger. (4, p. 182)

The above considerations establish the point of departure of the present experiment. Our aim is to determine the effects of degrees of hunger above zero hours and below three hours, on the strength of the behavior mediated by a previously learned Skinner-box, bar-depression habit of maximum strength. Three intensities of hunger were chosen for this initial attempt to determine the effects on behavior strength of weak degrees of drive: one-half, one, and two hours. The bar depression habit was built up to a point approximating maximum strength by allowing each animal 70 reinforcements. Habits of nearly maximum strength were established, in order to provide an optimal opportunity for the emergence of measurable differences in the effects of the very slight variations of the experimental variable (food privation). Since the present experiment was designed to fill in the, as yet, vaguely determined region of the Perin-Hull function between zero and three hours of hunger, a routine closely similar to Perin's was followed.

It should be repeated that the significance of the present experiment is not confined to its function as a further test of Perin's equations, and the associated Hullian assumptions about primary motivation. No previous attempt has been made to measure the possible facilitating effects of very

low drive intensities on the vigor of behavior. Most investigators tend to ignore or minimize the role of the presumably weak drives which may be present in various behavior situations, even if overshadowed by a seemingly dominant drive.[7] Notorious examples are such drive conditions as exploration, activity, escape from confinement; weak degrees of sex, thirst, or hunger. Indeed, in most experimental situations it is customary to control only one or a few drives, while all other drive factors which may be present are overlooked on the assumption that, relative to the dominant motivation, their effect on behavior is insignificant. If any inferences can be made from low intensities of hunger to low intensities of other drive modalities, the present experiment may be regarded as a preliminary test of this assumption.

Subjects and Apparatus

Both the present study and the earlier Koch-Daniel experiment were designed to measure empirically, values which Perin derived by theoretical extrapolation. Therefore, in both of these studies, it was necessary to duplicate, as closely as possible, the conditions of Perin's experiment.[8]

The data of the present experiment were obtained from 120 male, albino rats selected from the animal colony of the Duke Psychological Laboratory. These rats were approximately 90 days old at the beginning of the experiment, and they ranged in weight from 128 to 263 gm. Eight additional rats were used for preliminary standardization purposes, and 18 animals were discarded during the experiment for failure to satisfy the experimental conditions.

The apparatus consisted of two modified Skinner boxes of the variety used by Ellson (1), Williams (9), Perin (6, 7), and Koch and Daniel (4). Each box contains a 12 in. cubical, experimental compartment, enclosed in a large exterior shell which is packed with sound insulating material. The experimental compartment is ventilated by an air line which blows a gentle stream of compressed air into the apparatus, and it contains in one corner a bottle in which water is available throughout the experiment. On one wall of this compartment is a brass plate having a horizontal slot through which a horizontally oriented brass bar, 55 mm. in length, protrudes. Toward the bottom of the plate, and centered under the bar, is a small aperture through which pellets of food may fall into a food tray. During training, the apparatus was adjusted so that a pressure of 15 gm. or more, exerted on the bar, activated an electromagnetic food delivery device which released a single pellet of food into the food tray. At the same time, the bar depression activated: (1) a signal marker attached to a waxed paper polygraph, and (2) an impulse counter. In this way, a permanent record was obtained of the occurrence and latency of each response, and *E* could conveniently read the ordinal number of each response from the impulse counter.

The pellets were made from a mixture of finely ground Purina checkers, whole wheat flour and water, according to a procedure suggested by Hull and described by Siegel (8). Each pellet was a small, hard cylinder of this mixture, weighing approximately .075 gm.

During the extinction period, both the food delivery mechanism and the signal marker were disconnected. Thus, the only effect of bar-depression was the activation of the impulse counter, The duration of the extinction period, until the attainment of the five min. no response criterion. was timed with an electric timer.

---

[7] Due to the vagaries of current motivational terminology, expressions like 'weak drive,' 'low intensity drive,' etc., as used in this context, and throughout the present report, are likely to prove ambiguous. By a weak, or low intensity, drive, we do not mean one the facilitating effects of which on *behavior* are necessarily weak. Rather, we mean a drive which is low on the intensity continuum provided by some measure *independent* of the drive's effect on the vigor of behavior. About the only such measures of drive intensity available at the present time, are the various experimental operations in terms of which we roughly manipulate degree of primary motivation: e.g., food privation, water privation, temperature variation, etc. It should be clear that we are not entitled to assume that a drive which is 'weak' in this sense, will have a correspondingly weak effect on behavior strength, until this is empirically proven to be the case.

[8] At various points in our description of the apparatus and procedure of the present investigation. we will take the liberty of quoting freely from the Koch-Daniel report (4, pp. 171–173.)

Procedure

The two chief experimental conditions that our procedure was designed to realize were: (1) to reinforce the bar-depression response to a point approaching the limit of habit strength, under conditions of strong hunger, and (2) to extinguish the response at various brief time intervals after establishing complete hunger satiation. Measurements were made of the effects on resistance to extinction (behavior strength) of degrees of hunger resulting from three time intervals after satiation: one-half, one, and two hours. To this end, the 120 experimental animals were divided into three groups of 40 rats each: groups A, B, and C. The animals of all groups were trained to the limit of practice at the bar-depression reaction in the same way. Each group was then subjected to the extinction test at one of the three post-satiation time intervals—group A after one-half hour, group B after one hour, and group C after two hours. In achieving these conditions, an experimental routine similar to that of Perin (6), and Koch and Daniel (4), was followed. This routine may be divided into four phases: (1) establishing a feeding rhythm, (2) habituating the animals to the apparatus, (3) training at the bar-depression response, and (4) satiation and extinction.

1. *Feeding rhythm.*—All animals were fed moist food in their home cages for one-half hour on each of seven successive days. Each rat was fed at the same time of day at which he was later experimented upon. The food was a paste made of finely ground Purina checkers and whole wheat flour, mixed in the same proportion as in the dry pellets, and water. On days eight and nine, the rats were fed for one-half hour, at the customary time of day, in the apparatus. At these occasions, the food consisted of the standard pellets, a large quantity of which were placed in the food tray.

2. *Habituation.*—The purpose of habituation was to establish a connection in the animals between the characteristic thump of the food delivery device, and the tendency to secure and eat the resulting pellet of food. Thus, on the 10th day, each rat was fed 40 pellets, which were caused individually to appear in the food tray by *E* operating a hand switch mounted on top of the apparatus. When the animal was placed in the apparatus at the customary feeding time, three pellets were present in the food tray. Immediately after the rat consumed the third pellet, *E* released another pellet by closing the switch, and this was repeated until 20 pellets had been eaten. After the 20th pellet had been received, the rat was returned to its living cage for 15 min., and was then reintroduced into the apparatus where the same procedure was followed for another 20 pellets.

3. *Training.*—On the 11th day, the apparatus was adjusted so that it was necessary for the rats to depress the lever in order to secure a pellet of food, and the attempt was made to build up the bar-depression habit to a strength approaching the limit of learning. From Perin's curves for the 'number of reactions to produce extinction' as a function of 'number of reinforcement repetitions', it is apparent that 70 reinforcements eventuates in a habit strength close to the limit of learning. At 90 reinforcements, the largest number allowed any of the groups which serve as the basis for Perin's equations, the curves are virtually flat, and the acceleration from 70 to 90 is negligible.[9] Accordingly, the *E*s felt justified in assuming that 70 reinforcements were sufficient to generate habits of near maximum strength. Thus, each of the 120 rats was allowed 70 reinforcements.

The details of the training procedure follow. Before each rat was placed in the experimental compartment, a small quantity of moist food was smeared on the brass panel directly above the lever. This tended to orient the rat's exploratory movements toward the vicinity of the lever, so that, within a relatively short time, the rat would 'accidentally' depress the lever, and a pellet would appear in the tray. Three or four such accidental bar-depressions might occur, before the rat commenced to eat the food *immediately* after depressing the bar. *E* counted the first response which was immediately followed by the ingestion of the pellet as the *first reinforced response*, and, thereafter, the rat was allowed to remain in the box until the impulse counter registered 70 reinforcements.

4. *Satiation and extinction.*—On the 12th day, at the usual feeding time, each animal was satiated as rigorously as possible. After this, each animal was allowed to remain in its home

[9] Koch and Daniel (4) tested for a possible difference in the effects of 70 and 90 reinforcements, when reaction potential is measured at zero hunger, by allowing six of their 32 rats 90 reinforcements, and the remainder 70. Both of these subgroups gave precisely the same results.

cage for a time interval of one-half, one, or two hours, as the case may be, and the strength of any tendency to perform the bar-depression response was then tested by extinction to the five min. no response criterion.

Satiation was accomplished by placing a large container of moist food in the home cage of the animal at the customary time for experimentation. This food was left in the cage for a minimum time of one hour. If, at the end of about 55 min., the animal had stopped eating, *E* alternately offered it moist and dry food by hand until the food was refused. After this, the food container was left in the cage until there elapsed a two min. interval during which the animal did not eat. Although food was available to each animal for a minimum of 60 min., in no case did it require more than 70 min. for the animal to attain this two min. criterion of food refusal.

One-half hour after the end of the satiation procedure, each of the 40 rats of group A was placed in the apparatus for the extinction test. Each of the 40 animals of group B was subjected to the extinction procedure one hour after the end of satiation. The 40 animals of group C each underwent extinction two hours after the termination of satiation. During extinction, the food delivery device was disconnected, so that any depression of the bar would be followed neither by a pellet nor the thump of the food delivery mechanism.

In accordance with Perin's (and thus Koch and Daniel's) procedure, it was planned to measure the strength of the bar-depression behavior tendency (reaction potential) in three ways: "(1) the number of extinction . . . responses before the 5-minute criterion of no response; (2) the time required for extinction exclusive of the 5-minute criterion; (3) the latency or time required to make the first three extinction responses, a single latency being the temporal interval between the point in time at which an animal was placed in the experimental compartment and that at which it pushed the bar" (**6**, p. 98). In determining the latencies of the first three extinction responses, Perin tried to minimize the effects of the previous extinction on the latencies of the second and third responses, by returning the rats to their home cages for a one min. spontaneous recovery interval, after the first and second responses. In order to maintain a parallel routine, the same plan was adopted both in the Koch-Daniel experiment and the present one. All animals were given three extinction trials separated by the one min. recovery interval, even in cases where a rat satisfied the extinction criterion on the first or second trial. Thus, three 'latencies' were determined for every animal. *E* timed each latency with an electric timer. After the performance of the third response, the animal was left in the apparatus until attainment of the five min. extinction criterion. Total time preceding the criterion was measured by means of a second electric timer.

It is believed that the conditions of both the Koch-Daniel and the present study constitute an accurate reduplication of Perin's conditions in all essential respects. There were, however, three minor divergences from Perin's conditions, initially established in the Koch-Daniel study, and adhered to in the present experiment. (1) In the Perin experiment, the apparatus was ventilated by a suction line which sucked air out of the apparatus; in the present experiment air was blown into the apparatus. (2) Perin's pellets were made of a mixture of calves' meal and water; our own consisted of ground Purina checkers, whole wheat flour and water. (3) Perin's pellets weighed .050 gm. each; our pellets weighed .075 gm. each.

The difference in the direction of the air current could hardly affect our results in any significant way. Some possibility exists that the second and third differences indicated above could limit the comparability of our results with the corresponding values extrapolated by Perin, but only slightly, if at all. Hull assumes that the limit of the curve of habit strength is a positive growth function of both the quality and the quantity of the reinforcing agent (**2**, p. 133). If the nutritive value of our pellets differs from Perin's, it should follow that Perin's curve of reaction potential as a function of drive intensity, at the 70 reinforcement level, would have a somewhat different asymptote from that of the curve which would be generated by conditions like those of the present experiment. Any curve fitted to the reaction potential values measured in the present experiment might then not form an exact continuation of the curve based on Perin's values at higher intensities of hunger. Nevertheless, even if direct combination into a single curve of our values for weak hunger and Perin's for stronger hunger is not entirely legitimate, the general form of the two curves should be the same. In other words, the slight increment in the reaction potential values which might be produced by the possibly stronger reinforcing power of our pellets, would not interfere with comparison of the relative *rate* of rise of our curve, with the corresponding portion of Perin's.

Results

Fig. 1 provides a scatter diagram showing the main trend of the results. This diagram presents the numbers of responses preceding attainment of the five min. extinction criterion made by the individual animals in each of the experimental groups. A solid line has been drawn through the means of each distribution, while the medians are connected by an interrupted line. The mean values are 12.20 responses for the one-half hour hunger group (group A), 20.80 for the one hour hunger group (B), and 28.20 for the two hour

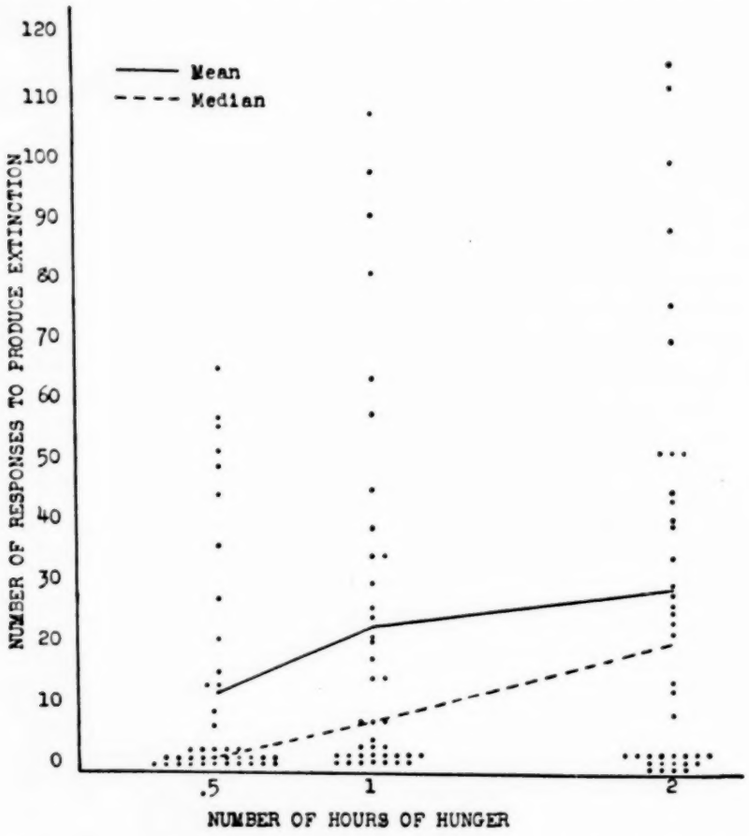

Fig. 1. The distributions of the number of extinction responses of the individual animals in each of the three experimental groups. The number of preceding reinforcements was held constant at 70 for all animals.

group (C). The medians are 1.00 response for group A, 6.00 responses for group B and 17.50 responses for group C. Both sets of values indicate behavior strength, as measured by number of extinction responses, to be a rapidly increasing function of hunger intensity between one-half and two hours. It should be observed, however, that the mean and median figures suggest trends of substantially different form. The curve conforming to the means shows definite deceleration in the interval between one and two hours of hunger, while the median trend from one-half to two hours approximates a straight line.

Each of the distributions is characterized by marked skewness, as is indicated by the high frequencies of zero scores, and the marked discrepancies between the mean and the median values. A glance at the distributions will also give the impression of great variability, the range being as wide as zero to 117 responses in the case of group C.

Table I presents average values for all three measures of behavior strength (reaction potential): number of responses preceding extinction, time preceding extinction, and the latencies of the first three

TABLE I

EFFECT OF THE NUMBER OF HOURS OF HUNGER AT THE TIME OF EXTINCTION ON THE MEASURES OF BEHAVIOR STRENGTH

Measure of Behavior Strength	Statistic	Number of Hours of Hunger (Reinforcements = 70 for All Groups)		
		Group A One-Half Hour	Group B One Hour	Group C Two Hours
(1) Number of Extinction Responses Preceding the Five Min. Criterion	Mean	12.20	20.80	28.20
	Median	1.00	6.00	17.50
	σ_{mdn}	3.87	5.80	2.08
(2) Time Preceding Extinction (Min.)	Mean	6.64	9.60	10.86
	Median	3.67	7.50	8.59
	σ_{mdn}	1.39	1.67	1.78
(3) Latency of Initial Extinction Responses (Min.)*	Median	4.56	2.70	2.09

* Since a number of animals did not make any responses within the five min. which were allowed, several indeterminate latencies were obtained, thus making computation of the means impossible. Consequently, no σ_m was available from which to compute the σ_{mdn}. (Cf. Perin, 6, p. 100, footnote to Table III.)

extinction responses. Both means and medians are given for the number of extinction responses, and extinction time. Only medians are included for the latency scores.[10] Any means computed from the latency data would necessarily be ambiguous, because each of the groups contained a large number of animals who did not make a single extinction response. As Perin (6, p. 100) and Koch and Daniel (4, pp. 174–175) have already pointed out, the 'latency' scores for such animals are indeterminate, having some value between five min. and infinity. Any mean based even partially on such values must itself be indeterminate.

Fig. 2 presents graphs of the relations between the average behavior strength values and the three hunger intensities of the present experiment. Separate graphs are given for each of the three meas-

[10] It will be recalled that the latencies for the initial *three* extinction responses were determined for each animal. This gave 120 latency scores for each of the experimental groups of 40 rats. The median latency for a given group was based on the combined distribution of all 120 scores.

ures of behavior strength. These graphs consist of lines drawn through points plotted from the data in Table I. Continuous lines represent means, and broken lines, medians. In order to extend the graphs to the zero level of hunger, mean and median behavior strength values derived by Koch and Daniel (4), under conditions

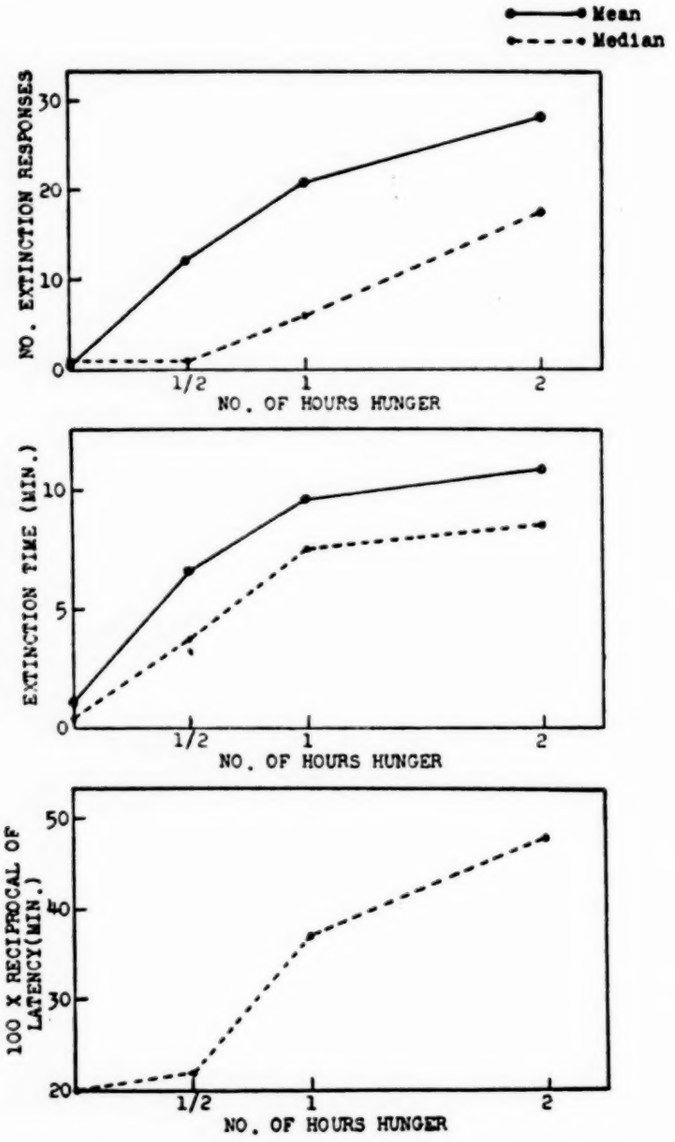

FIG. 2. Curves showing each of the three measures of behavior strength as a function of durations of food deprivation ranging from zero to two hours. The values for 'zero' hunger (complete satiation) were taken from the Koch and Daniel experiment (4). The values for one-half hour, one hour and two hours of hunger were plotted from the data provided by groups A, B, and C of the present experiment. The 120 rats of the present experiment (40 in each group) were subjected to extinction after 70 reinforcements; the Koch-Daniel data represent combined averages for 26 rats extinguished after 70 reinforcements, and six rats extinguished after 90 reinforcements. In order to facilitate comparison of the latency curves with the others, the *reciprocals* of the median latencies were plotted, instead of the obtained medians. The median initial latency obtained by Koch and Daniel for zero hunger was indeterminate, having some value between five min. and ∞ (cf. Table III and text). Since five min. represents the maximum latency obtainable under the conditions of these experiments, it was considered legitimate to plot 100 $\times$ the reciprocal of five min. (i.e., 20) at the *origin* of the latency curve.

comparable to those of the present study, are also represented in Fig. 2.

From both Table I and Fig. 2, it is clear that all three measures of behavior strength agree in showing a rapid increase from one-half to two hours of hunger under the present experimental conditions.[11] This is the only respect in which the trend of all the obtained averages can be said to agree. We have already referred to the discrepancy in the form of the curves suggested by the mean and median values for the number of extinction responses. Although the mean and the median extinction time values are somewhat more consistent in trend than are the corresponding averages for the number of extinction responses, both of the extinction time averages generate somewhat different curves from either of the curves for the number of extinction responses. Both the mean and the median extinction time curves show marked deceleration in the one to two hour interval. This same tendency for the rate of increase in behavior strength to be distinctly slower in the one to two hour interval is also suggested by the median latency curve. Because of the skewness of the distributions of extinction data deriving from their Skinner-box procedure, Perin (**6**) and Hull (**2**) have adopted the convention of treating such data in terms of medians rather than means. Nevertheless, even when we restrict scrutiny to the form of the three median curves, we find that the straight line function defined by the extinction response values (from one-half to two hours of hunger) diverges radically from the negatively accelerated trend of the extinction time, and initial latency, values.

TABLE II

RELIABILITIES OF THE DIFFERENCES BETWEEN THE MEASURES OF BEHAVIOR STRENGTH: CRITICAL RATIOS (DIFF./$\sigma_{diff.}$) OF MEANS AND MEDIANS FOR PAIRED GROUPS

Paired Hunger Groups	Statistic	Measure of Behavior Strength	
		C.R. of No. of Extinction Responses	C.R. of Extinction Times
Group A vs. Group C (½ hr. vs. 2 hrs.)	Mean	4.61	4.58
	Median	3.76	2.22
Group B vs. Group C (1 hr. vs. 2 hrs.)	Mean	1.50	.65
	Median	1.87	.45
Group A vs. Group B (½ hr. vs. 1 hr.)	Mean	1.54	1.71
	Median	.72	1.76

[11] Attention should be drawn to the fact that it is not the obtained latency, but the reciprocal of the latency, which serves as a measure of behavior strength. The *decreasing* median latency values in Table I actually represent *increasing* behavior strengths as measured by initial latencies. In order to make the curves in Fig. 3 more readily comparable, the reciprocals of the median latencies were plotted instead of the obtained medians.

Table II presents critical ratios indicating the reliability of the differences between average behavior strength measures for the three possible pairings of the experimental groups. Critical ratios could not be computed for the median latencies, because of the method used in determining the latencies (cf. Perin, **6**, p. 100, footnote to Table III).

It is apparent in Table II that the differences between the average values of groups A and C (the one-half hour vs. the two hour group)

TABLE III

EMPIRICALLY OBTAINED MEDIAN BEHAVIOR STRENGTH VALUES FOR INTENSITIES OF HUNGER FROM ZERO TO TWO HOURS COMPARED WITH CORRESPONDING THEORETICAL VALUES DERIVED FROM PERIN'S EQUATIONS *

Measure of Behavior Strength (Reaction Potential)	Source of Data	Conditions Constitutive of the Behavior Strength (Reaction Potential)			
		Reinforcements = 70			
		Hunger = 0 Hours	Hunger = ½ Hour	Hunger = 1 Hour	Hunger = 2 Hours
Median Number of Extinction Responses Preceding the Five Min. Criterion	Obtained in Koch-Daniel and Present Experiment	1.00	1.00	6.00	17.50
	Derived from Perin's Theoretical Equation	16.27	16.79	17.33	18.45
Median Time Preceding Extinction (Min.)	Obtained in Koch-Daniel and Present Experiment	.36	3.67	7.50	8.59
	Derived from Perin's Theoretical Equation	9.22	9.22	9.23	9.27
Median Latency of Initial Extinction Responses (Min.)	Obtained in Koch-Daniel and Present Experiment	Indeterminate (5 min.—∞)	4.56	2.70	2.09
	Derived from Perin's Theoretical Equation	.54	.53	.53	.51

* The theoretical values were derived from Perin's equations (**6**, pp. 108–111) for the measures of behavior strength as a function of number of reinforcements ('R'), and number of hours of hunger ('h'). These equations are:

(1) $$N = 21.45 \times 10^{.0222h}(1 - 10^{-.018R}) - 4.$$

(2) $$T = 11.0 \times 10^{.00053h^2}\left[1 - 10^{\frac{R}{16}\log\left(1-.867\times10^{-.00025h^2}\right)}\right] - 1.5.$$

(3) $$L = \frac{(.43 - .01h)R + (7.22 - .275h)}{R - 1}.$$

N stands for the number of extinction responses, T for time preceding extinction, and L for latency. Since the curves were fitted to median values throughout, all solutions of these equations for N, T, and L represent medians. For this reason, it was impossible to compute theoretical *means* for Perin. The theoretical medians attributed to Perin in the table were arrived at by solving each of the equations after setting R and h at the indicated values.

show satisfactory reliability with respect to both of the measures of behavior strength. The differences between both the mean and median number of extinction responses for groups A and C are statistically significant at the one percent level. The mean extinction time difference, as between the same two groups, is also significant at the one percent level. The difference between the median extinction times fro groups A and C, is significant at the five percent level.

The differences which result from the other two pairings of the groups are not statistically significant. Nevertheless, for such slight variations of the experimental variable (food privation), the *consistency* of the increasing trends, in the case of all measures of behavior strength, is impressive.

Table III presents a comparison of the median behavior strength values obtained in the Koch-Daniel and the present experiment, with the corresponding values predicted by Perin's equations. The disagreement between the trends of the empirically determined and the theoretically derived values is striking. All three of the empirically obtained measures show a rapid and progressive build-up, from a near zero behavior strength at zero hunger to a substantial value at two hours' hunger. Perin's extrapolated values, for all three behavior strength measures, are relatively high at zero hunger, and show only minute increments as hunger is increased to the two hour level. The obtained median number of extinction responses ranges from 1.00 at zero hunger to 17.50 at two hours' hunger, while the range of the corresponding values implied by Perin's equations is from 16.27 to 18.45 responses. The measured median extinction time values are .36 min. at zero hunger, and 8.59 min. at two hours' hunger; Perin's corresponding extrapolations are 9.22 min. and 9.27 min. The range of the empirically determined, median initial latencies is from 5 min. to ∞ (indeterminate) at zero hunger to 2.09 min. at two hours; the range of Perin's corresponding theoretical values is .54 min. to .51 min.

The wide discrepancy in the ranges of the obtained, as against the theoretical, behavior strength values should not obscure another trend suggested by Table III. As hunger intensity increases from zero to two hours, the *amount* of disagreement between the measured behavior strength values, and those yielded by Perin's equations progressively decreases. Although the disagreement is still marked at the one hour hunger level, by two hours of hunger the disagreement —at least in the case of two of the measures of behavior strength— becomes much smaller. These findings are consistent with the possibility, already suggested by Koch and Daniel (**5**, p. 170), that the gross inaccuracy of the Perin-Hull function at very weak degrees of hunger may decrease at higher hunger intensities.

An impression of the relation between behavior strength and hunger intensity for a habit of near maximum strength, so far as this relation has yet been explored, may be derived from the curve in Fig. 3. This curve represents one of the measures of behavior strength—median number of extinction responses—as a function of intensities of hunger ranging from zero to 22 hours, when preceding reinforcements are held constant at 70. The value for zero hunger is plotted from the results of the Koch-Daniel study. Data points for one-half, one, and two hours of hunger are provided by the present experiment. The behavior strength values for degrees of hunger ranging from three to 22 hours are derived from Perin's equations.

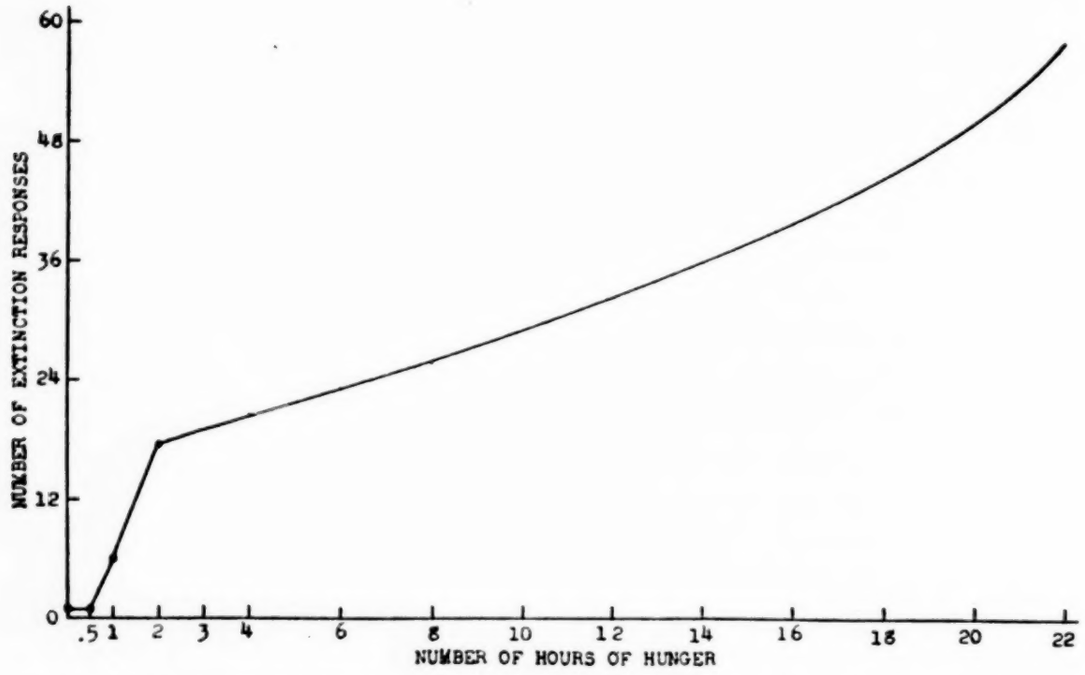

FIG. 3. Diagrammatic curve showing the median number of extinction responses as a function of the number of hours' hunger from zero to 22 hours, for a bar-depression habit at the 70 reinforcement level. The solid circles represent empirically obtained medians deriving from the Koch-Daniel and the present experiment. The portion of the curve after two hours is a smoothed line drawn through values derived from Perin's equation for the number of extinction responses (equation 1, footnote to Table III). Since the obtained medians for the zero to two-hour interval may represent points on a curve having a different asymptote than does Perin's curve, Fig. 3 must be regarded as a crude diagram intended to depict the relationship as adumbrated by the available evidence.

The curve of Fig. 3 must be regarded as highly tentative. This is not only because of the sparseness of the data on which the portion of the curve derived from Perin's equation (between three and 22 hours' hunger) is based. We must also recall, as stated above, that there may be a difference in the nutritive value, and thus the reinforcing power, of the pellets used in both the Koch-Daniel and the present experiments, as against those used by Perin. There is thus

some possibility that the values obtained by Koch and Daniel, and the present investigators, may represent points on a curve the asymptote of which does not coincide with the asymptote of Perin's curve. The combination of the two sets of values into a single curve may only be justified as a device for diagramming the relative rates of rise of different intervals of the relationship.

Inspection of Fig. 3 clearly reveals a much more rapid rate of increase of behavior strength between zero and two hours' hunger than at any higher interval of the curve.[12] As predicted by Koch and Daniel (4, p. 182), a small increase in the number of hours of food privation at low intensities of hunger, seems to exert a relatively greater effect on the vigor of behavior than similar increases at higher intensities of hunger.

A partial check on the degree to which the present investigation succeeded in reproducing the conditions of Perin's experiment may be made by comparing the initial learning efficiency of the groups used in the two studies. The only measure of initial learning ability reported by Perin is the time required for the performance of the first five reinforced responses. Accordingly, Table IV presents the mean and median times required for the first five reinforced responses by each of the groups of this experiment, averages of these values for all three of the present experimental groups taken together, and averages of the initial learning times for all eight of Perin's groups. It is obvious from Table IV that the combined averages for the

TABLE IV

COMPARISON OF THE AVERAGE INITIAL LEARNING ABILITY OF THE PRESENT EXPERIMENTAL GROUPS WITH THE AVERAGE INITIAL LEARNING ABILITY OF PERIN'S SUBGROUPS

Statistic	Min. Needed to Secure First Five Pellets				
	Group A	Group B	Group C	Average of Groups A, B, C	Average of Perin's Groups
Mean	2.21	2.62	2.84	2.56	2.28
Median	1.99	2.18	2.34	2.17	2.03
σ_{mdn}	.15	.33	.28	.25	.21

groups of the two experiments are in close agreement. The critical ratio of the averaged medians of Perin's eight groups and the averaged medians of the groups of the present experiment is only .42, indicating that the probability is very great that the rats of the two ex-

[12] This statement holds, of course, only for that portion of the total relationship between hunger intensity and behavior strength which has been empirically explored. The limiting value on which Perin's curve is based is 23 hours. Since the portion of the present curve based on Perin's data shows definite positive acceleration, it is possible that intervals of the relationship after 23 hours might turn out to show higher rates of rise than the zero-two hours interval.

periments are drawn from the same population with respect to the measure of initial learning efficiency.

Discussion

The present experiment is significant in two main ways. It provides an initial, tentative estimate of the effects of weak degrees of motivation on behavior strength, and it serves as a test of the Perin-Hull function for low intensities of drive (between zero and two hours). We will discuss the results from each of these points of view, in turn.

Very little experimental information has been available on the *quantitative* relations between drive intensity and the vigor of behavior. Nevertheless, it has been customary for both theorists and experimenters to assume a close correspondence between the apparent intensity of a drive and the magnitude of its energizing effects on behavior. Indeed, the distinction between the *strength of a drive*, as measured by the standard experimental operations for manipulating drive states (e.g., food privation), and the *strength of the drive's effect on behavior* has seldom been clearly recognized. Thus, expressions like 'drive intensity' have become systematically ambiguous, sometimes referring to the status of a drive on the continuum provided by some measure independent of the drive's effect on behavior, sometimes referring to the magnitude of the drive's facilitating effect on behavior, and frequently referring to both at once. This confusion has been responsible for the uncritical, but almost universal, tendency to assume that a drive which is 'weak' in the first sense, is correspondingly weak in its effects on behavior.

Experimenters have found this last assumption to be a great convenience. Drive states, particularly 'weak' ones, are notoriously difficult to control. If one conducts an experiment in which it is necessary to vary a single drive and eliminate or hold constant all others, it is comforting to believe that an hour of hunger, a slight temperature change, a weak sex drive, or a feeble activity drive will make no appreciable difference. The results of the present experiment call clearly into question any such assumption. At least in the case of hunger, degrees of food privation as slight as one-half, one, and two hours, turn out to have a considerable influence on the vigor of behavior.

Whether these results can be generalized to comparably low intensities of other drives is by no means certain. But, in the absence of further evidence, we can hardly conclude that the reverse is the case. The responsibility for experimentally testing the effects on behavior strength of low intensities of other drives must, in any event, be squarely faced.

The immediate implications of the present experiment, as a test of the Perin-Hull function at near maximum habit strength and weak intensities of drive, are unambiguous. For hunger intensities of one half and one hour, and a habit strength of 70 reinforcements, the Perin-Hull function has been shown to be grossly inaccurate. At the two hour hunger level, the theoretical and empirical values fall more closely into line, but even here the agreement is not complete. If we now recall, form the Koch-Daniel experiment, that Perin's equations are markedly in error at zero hunger as well, we can conclude that for intensities of hunger from zero to about two hours, preceding reinforcements being held constant at 70, the Perin-Hull function does not accord with experimental measurement.

The above statement, it should be noted, asserts that the Perin-Hull function proves inaccurate for weak degrees of hunger, when habit strength is at the near maximum value of *70 reinforcements*. Does the function remain erroneous for habit strength values below 70 reinforcements? The logic of Perin's curve fitting procedure implies that this should be the case. The Perin-Hull equations are derived on the assumption that the *general form* of the relation between reaction potential and hunger intensity will remain constant at every level of habit strength (i.e., number of past reinforcements). The only empirical data which Perin used for determining the relation between reaction potential and drive intensity consisted of four groups of rats extinguished after 16 reinforcements. In deriving the equations, it was assumed that the drive intensity curve would retain the same form at *every* number of reinforcements, but that the asymptote of each curve would vary with the number of reinforcements (according to a positive growth function). If we hold fast to this assumption, the marked inaccuracy of the Perin-Hull function at weak hunger and 70 reinforcements, may be extrapolated to corresponding degrees of hunger and any lower number of reinforcements.

As we just indicated, the Perin experiment includes four groups extinguished after 16 reinforcements, but under varied strengths of hunger. The four degrees of hunger were: one, three, 16 and 23 hours. Since the present experiment also includes a one-hour hunger group, it should prove instructive to compare the reaction potential values obtained in the present experiment at one hour of hunger and 70 reinforcements, with the results of the group extinguished by Perin at the same level of hunger and 16 reinforcements. At *16* reinforcements and *one hour* of hunger, Perin's group yielded a median number of extinction responses of 6.00 (**6**, p. 103). At *70* reinforcements and *one hour* of hunger, the present experiment also yields a value of 6.00 extinction responses. This means that two independ-

ent experiments which seek to determine reaction potential at the same hunger intensity (one hour), but at grossly different levels of habit strength (16 *vs.* 70 reinforcements), arrive at the same value.[13] According to Perin's equation, the median number of extinction responses at 70 reinforcements and one hour of hunger should be not 6.00, but as high as 16.27.

The failure of the present experiment to find a substantially higher reaction potential, at one hour of hunger and 70 reinforcements, than did Perin, at the same hunger strength and 16 reinforcements, is difficult to explain. It may, however, be ascribed to any, some, or all of the following possibilities: (1) One of the measurements may have failed adequately to realize the experimental conditions, and thus yielded a false value. (2) One or both of the values may have been distorted by a gross sampling error. (3) It may not be legitimate to assume that the relation between reaction potential and drive intensity remains constant in form, at every number of reinforcements (habit strength).

The first possibility can only be decisively evaluated by further experiment. So far as the writers can determine, the present experiment is an accurate reduplication of Perin's experiment. If the divergence of our one-hour value from theoretical expectation can be interpreted as an indication of inadequate reproduction of Perin's conditions, then the approximate agreement with theoretical expectation of our two-hour median must be accepted as an indication of the reverse. Nevertheless, it would be desirable to re-measure the reaction potential values at one hour hunger and 16 reinforcements, and one hour and 70 reinforcements, in the same laboratory.

The possibility of sampling error—at least partially—accounting for the identity of the one hour hunger medians at both 16 and 70 reinforcements, is a very real one. We have already pointed to, and exhibited in our results, the marked variability of the type of Skinner-box data recorded in the experiments under consideration. This variability is so pronounced that sizable sampling errors seem almost unavoidable. In the present case, the role of sampling error may be partly evaluated by comparing, in addition to the medians, the obtained means, and the sigma medians. A very

[13] Perin's theoretical equation yields the even more paradoxical median value of 7.06 extinction responses for conditions of 16 reinforcements and one hour of hunger.

The comparison in the text is restricted to only one of the measures of behavior strength: the number of extinction responses. The situation becomes even worse when we make the comparison in terms of a second measure, initial latency. Perin's obtained, median initial latency, for his *16* reinforcement and one hour hunger group, was .95 min. Our own median for the group extinguished at *70* reinforcements and one hour of hunger was 2.70 min. Recalling that behavior strength is the *reciprocal* of the initial latencies, it is clear that here we have the paradox of a much *lower* behavior strength at *70* reinforcements and one hour of hunger than at 16 reinforcements and the same level of hunger. The remaining measure of behavior strength—time preceding extinction—gives a difference in the theoretically expected direction (Perin's obtained median being 4.45 min. and our own, 7.50 min.), but our own median is distinctly lower than the median for 70 reinforcements and an hour of hunger predicted by Perin's equation (cf. Table III).

different picture emerges when the means are compared instead of the medians. Perin's median for one hour of hunger and 16 reinforcements, and our own median for one hour of hunger and 70 reinforcements, are both 6.00 extinction responses; the corresponding *means*, however, are 9.10 responses at the 16 reinforcement level, and 20.80 responses at the 70 reinforcement level. In other words, when *means* are used as the average for the number of extinction responses, a difference in the expected direction between the reaction potential at 16 and at 70 reinforcements clearly emerges. A comparison of the obtained sigma medians—the only direct measure of dispersion which Perin reports—suggests that variability was substantially greater in the present one hour hunger and 70 reinforcement data (5.80 responses) than in Perin's one hour hunger and 16 reinforcement data (2.01 responses).[4] These circumstances make it highly probable that the paradoxical identity of the median number of extinction responses at 16 and 70 reinforcements is due largely to errors of sampling.

The third possibility—to the effect that reaction potential may *not* bear a constant relation to hunger intensity at every number of reinforcements—cannot be discarded without experimental test. It is, of course, convenient to make this assumption, but the convenient and the true do not always coincide. If the identity of the one hour hunger medians at the two different levels of reinforcement should turn out to be due, in any measure, to this factor, then the entire Perin-Hull function would break down. Indeed, in such a case the very possibility of reducing reaction potential to a unified joint function of drive intensity and habit strength would disappear. It would then be necessary to derive separate relationships between reaction potential and drive intensity for an indeterminate number of different habit strengths. Even if the paradox under consideration can be entirely accounted for in terms of sampling errors—which is probably the case—it would still be desirable to test the assumption of a constancy in form of the hunger intensity curves at different reinforcements levels.

The preceding discussion has once more brought us up against a point which we took pains to emphasize above in the presentation of our results: the pronounced variability of the data yielded by the Skinner-box technique used by Perin and the present investigators. If this variability permits sampling errors which could produce the possible distortion of averages that we just considered, then it becomes hard to maintain confidence in the *quantitative precision of most of the values derived in the entire series of experiments.* When we add to the variability, the closely related fact of the marked skewness

[14] The admittedly greater variability in the present one hour hunger data may lead to the inference that this reflects inadequacy or instability of our experimental procedure. Such an inference would not be fair. In our two hours' hunger and 70 reinforcement group, which was subjected to precisely the same routine, the sigma median was 2.08 responses. One of the groups in the Perin experiment was tested at three hours' hunger and 70 reinforcements. Here the sigma median was 4.88 responses. One might infer with equal cogency that in this case *Perin's* procedure stands indicted of laxity. Short of the unscrupulous selection of cases, there seems to be no way of controlling the variability of the data gained through procedures of the sort used by Perin and the present investigators.

of all of the distributions, the picture becomes very depressing indeed. One consequence of the skewness is the wide divergence between the mean and median values. As we have already seen, in the present experiment, this discrepancy was such that, should we plot the reaction potential curves in means rather than medians, not only would the absolute values change, *but so would the actual form of the curves.*

It is difficult to avoid concluding that, before further attempts are made to apply Skinner-box methods of the type used in the present series of studies to the discovery of *quantitative* behavior laws, the problem of minimizing both variability and skewness of data must be given careful thought. We might suggest, in passing, that the problem may fruitfully be approached from two directions: statistically, and from the incidence of experimental procedure. Statistically, it should be possible to compensate for both variability and skewness by adopting conventions for the selection of cases that would give less weight in the averages to the many zero, and the few extremely high scoring, animals. It seems plausible to assume that such animals are not being adequately affected by the experimental variable (hunger intensity); an animal who scores zero extinction responses, under one, or two, or 22 hours of hunger, may be satisfying the experimental conditions no more adequately than a rat who fails to learn the bar-depression response, and is thus discarded. Experimentally, many modifications of procedure, which could conceivably reduce variability, might be introduced. One such change might be the elimination of the three spontaneous recovery intervals during the extinction test. Again, Walker has apparently already demonstrated that a reduction of the area of free movement, within the Skinner-box, decreases variability.[15]

In general, the results of the present experiment permit us to conclude that, for a Skinner-box bar depression habit of near maximum strength, increases of hunger intensity form zero to two hours produce a rapid increase of behavior strength (reaction potential). Further, if we assume the approximate validity of Perin's theoretical behavior strength values for higher intensities of hunger, the rate of increase of behavior strength as a function of hunger intensity is greater within the zero to two hour interval, than it is for any higher interval of the curve.[16] Because of the variability and the skewness of the data, it would be rash to assert that the median behavior strength values reported in the results are quantitatively exact, or that they represent even the true *form* of the relationship between behavior strength and hunger intensity.

We must remember that the Perin-Hull function figures promi-

[15] Walker, E. L. An investigation of factors producing variability in extinction of the conditioned response established in white rats using the Skinner-box technique. (Unpublished manuscript cited by Kendler, 3)

[16] Twenty-three hours is the highest hunger intensity for which the Perin-Hull function was determined. Thus, the statement in the text holds only up to the 'known' limit of 23 hours.

nently in that part of Hull's current theory of behavior devoted to the analysis of 'primary motivation' (**2**, pp. 226–257). So far as can be determined from the present formulation of the theory, the Perin-Hull function appears to provide the chief empirical basis for the key postulate on primary motivation, 'postulate 7,' and perhaps for certain other assumptions—both expressed and unexpressed—which are presupposed in the treatment of primary motivation.[17] The Koch-Daniel experiment demonstrated the inaccuracy of this function for zero hunger and all habit strengths. The present experiment extends the known zone of inaccuracy from zero to some point near two hours of hunger, for habits of approximately maximum strength. If we accept Perin's assumption of a constancy in form of the hunger intensity curves at all numbers of reinforcements, the inaccuracy of the function for weak degrees of hunger ramifies throughout every level of reinforcement. All derivations within

[17] Although postulate 7 seems to be the *major* assumption in Hull's present theory of primary motivation which is based on Perin's equations, other assumptions such as 'postulate 6' (**2**, p. 253), 'major corollary II' (**2**, p. 253), and certain of the assumptions governing Hull's "detailed illustration . . . of the theory of motivation" (**2**, pp. 242–247), seem largely to be guided by the Perin function.

Postulate 7, as formulated in *Principles of behavior*, represents a *verbal* restatement of what is believed to be one of the most important characteristics of Perin's equations: namely, that reaction potential is a *multiplicative* function of habit strength and drive intensity. An indefinite number of differing *quantitative* relationships could conceivably be compatible with this verbal formulation. Thus, our results are not necessarily in conflict with the *verbal* form of postulate 7 (cf. **4**, p. 182). Nevertheless, since most of the 'derivations' which are apparently mediated by postulate 7 seem to presuppose some or all of the quantitative details of the Perin equations, it seems legitimate to regard Perin's equations as the *quantitative* formulation of postulate 7, at least as that postulate *functions* within the theory put forward in Hull's book. This identification of the Perin-Hull function with the *quantitative* formulation of postulate 7 was explicitly made in the report of the Koch-Daniel experiment (**4**, p. 170). When the Perin-Hull function is equated with the quantitative formulation of postulate 7, it is obvious that any empirically demonstrated defect in the Perin equations must also represent a defect in postulate 7. In this sense, the results of both the Koch-Daniel and the present experiment are distinctly in confluct with the *quantitative* formulation of postulate 7.

C. B. Woodbury, however, has published a note (**10**) challenging this interpretation of the Koch-Daniel experiment. His argument has been analyzed in a reply by Koch and Daniel (**5**). Space forbids a detailed re-examination of Woodbury's contentions. Let it suffice to say, however, that the 'interpretation' suggested by Woodbury would enable Hull's 'theory' to generate directly opposed theorems with regard to certain of the phenomena of primary motivation. As Koch and Daniel have shown, Woodbury's arguments, themselves vague and self-contradictory, bring to light a basic ambiguity in Hull's theory of motivation itself.

In view of this confusing situation, perhaps the best way to state the bearing of our results on Hull's theory of motivation is as follows: Hull's theory of motivation, as interpreted and used by Woodbury, turns out to be such that it is impossible to specify, at present, *what its major assumptions are*. Therefore, if we abandon the apparently futile task of locating the specific assumptions in the 'theory' which are incompatible with our results, we can conclude that our results are in conflict with *whatever use is made in Hull's theory of the Perin-Hull function*, when set at drive intensities equivalent to zero to about two hours of hunger. Whether this formulation leaves the theory on any firmer ground than does the interpretation in terms of postulate 7 is, of course, questionable. For a detailed justification of the foregoing points, the reader should consult Woodbury's note, and the Koch-Daniel reply.

Hull's system involving the present form of the Perin-Hull equations, when set at intensities of hunger from zero to about two hours, will necessarily be incorrect.

The above paragraph summarizes the immediate implications for the Perin-Hull function, and the theory of behavior in which it is incorporated, of the values obtained in the Koch-Daniel and the present experiment. But evidence has been adduced which throws serious doubt on the ability of the Perin-type, Skinner-box method to yield quantitatively reliable behavior strength values. To the writers, this evidence is sufficiently compelling to place the *entire* Perin-Hull function under tentative suspicion. If, nevertheless, it is felt fruitful to retain the Perin-Hull function in a form resembling its original one, then it must be recast in the light of the present results.

Summary

This experiment was designed to measure the effect of three weak intensities of the hunger drive on the vigor of performance (behavior strength) of a simple habit, already learned to the limit of practice. The experimental conditions were realized in the following way. One hundred and twenty male, albino rats, divided into three groups of 40 animals each, were allowed 70 reinforced bar-depression responses in a modified Skinner box. All animals were motivated, during the acquisition of this habit, by 23 hours of hunger. On the following day, hunger in each rat was satiated according to a standard procedure, and, after a brief time interval which varied for the three experimental groups, the bar depression habit was subjected to extinction. The intervals of food privation between satiation and extinction were one-half hour for one of the groups, one hour for a second group, and two hours for the third. The strength of the behavior mediated by the bar-depression habit, under the conditions of the extinction test, was measured in three ways: (1) number of bar-depressions preceding a five minute no response interval, (2) time preceding the five min. extinction criterion, and (3) latencies of the first three extinction responses.

The results show that intensities of hunger resulting from amounts of food privation as slight as one half to two hours, exert a definite facilitating effect on the strength of the learned behavior. Between one half and two hours of hunger, all three of the measures of behavior strength show a rapid and progressive increase.

The experiment is significant in two chief ways:

1. It represents one of the first quantitatively oriented attempts to measure the effects of very weak degrees of a primary drive, as

measured by the standard operations for varying such a drive, on behavior strength. Investigators have often assumed that drives which are weak according to criteria independent of the drive's effect on behavior (e.g., food privation, water privation, etc.) are correspondingly weak in their actual effects on behavior. But the present experiment has shown that intensities of hunger as 'weak' as one half to two hours may exert a substantial influence on the vigor of behavior.

2. The present experiment functions as a direct test of certain implications of Perin's equations (**6**) which express the results of an elaborate, curve-fitting experiment intended to determine the joint functional interdependence of three of Hull's chief constructs: 'drive intensity' (e.g., hours of hunger), 'habit strength' (number of reinforcements), and 'reaction potential' (behavior strength). These equations serve as the main empirical basis of what seems to be the major postulate relevant to Hull's present analysis of 'primary motivation,' postulate 7, and possibly of other assumptions in Hull's treatment of motivation. A previous experiment by Koch and Daniel had shown that this Perin-Hull function yields highly erroneous reaction potential values, under conditions of zero drive and any given habit strength. The results of the present experiment indicate that the Perin-Hull function remains markedly out of line with experimental measurement at the one half and one hour levels of hunger. At two hours of hunger, the theoretically predicted and the obtained values show fair, but not complete, agreement in the case of two of the measures of reaction potential (number of extinction responses and extinction time), but the divergence remains marked for the third measure—the initial latencies.

The paradoxical agreement between behavior strength values yielded by one of the present groups and one of Perin's, both extinguished at one hour of hunger but after grossly different numbers of reinforcements—together with the evident variability of Perin's and the present distributions—suggests that the Perin-type, Skinner-box procedure may be inadequate for the discovery of *quantitatively* precise behavior laws. Attention might fruitfully be directed toward modifying the present procedure so as to reduce the variability and skewness of the data.

If we accept as approximately correct at least the *general form* of the Perin-Hull function, as it holds for hunger intensities from about three to 23 hours, a major finding of the present experiment may be stated as follows: The rate of increase of behavior strength as a function of hunger intensity (for habits of near maximum strength) is relatively greater from zero to some point between two and three

hours of hunger than at any higher interval of the hunger curve up to the limiting, previously explored value of 23 hours.

(Manuscript received July 30, 1947)

References

1. Ellson, D. G. Quantitative studies of the interaction of simple habits: I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
2. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
3. Kendler, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp. Psychol.*, 1945, **35**, 96–109.
4. Koch, S., & Daniel, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, **35**, 167–187.
5. Koch, S., & Daniel, W. J. Form and content in hypothetico-deductive systems: a reply to Woodbury. *J. Psychol.*, 1947, **24**, 237–246.
6. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
7. Perin, C. T. The effect of delayed reinforcement upon the differentiation of bar responses in white rats. *J. exp. Psychol.*, 1943, **32**, 95–109.
8. Siegel, P. A note on the construction of Skinner box pellets. *J. comp. Psychol.*, 1944, **37**, 233–235.
9. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–521.
10. Woodbury, C. B. The meaning of Koch and Daniel's results for Hull's theory. *J. exp. Psychol.*, 1947, **37**, 194–196.

THE EFFECT OF SUCCESS AND FAILURE UPON THE ATTRACTIVENESS OF ACTIVITIES AS A FUNCTION OF EXPERIENCE, EXPECTATION, AND NEED *

BY MILDRED E. GEBHARD

University of Pennsylvania

INTRODUCTION

It is a common assumption that success in activities increases the attractiveness of those activities while failure decreases their attractiveness. The results of an experimental study by Cartwright (2) indicate, however, that successful completion of an activity led some *S*s to consider it more attractive, but that others found it less attractive.[1] Likewise, with failure the attractiveness of the failed activity decreased for some *S*s but increased for others. Cartwright, himself, on the basis of post-experimental interviews with his *S*s, suggests that uncontrolled variation in the experience and expectation of the *S*s may have operated to produce these changes but he presents no conclusive evidence of this.

The present experiment attempts to study further the changes in the attractiveness of activities resulting from variations in the experience and expectation of success and failure. In addition, a third variable, the strength of need is employed. Each of these three variables will be considered in detail in the discussion which follows.

1. *Experience.*—One of the basic hypotheses advanced in this study is that the attractiveness [2] of an activity, in general, will rise with the experience of success and fall with the experience of failure. This prediction receives partial support from several studies in which *S*s choosing between a pair of tasks with which they had had contrasting experience stated a

* This paper is a dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy, in the Department of Psychology of the University of Pennsylvania. The writer wishes to express her sincere appreciation to Professor Francis W. Irwin for his many helpful suggestions and stimulating criticisms offered throughout the course of the experimental investigation. The writer is similarly grateful for guidance received from the Departmental Group Committee. To the students and faculty of Beaver College, and especially to Professor Dallas Buzby, the writer is indebted for gracious cooperation in the actual completion of the experimental work.

[1] The *S*s were told that they had done 'all right,' that their time was about average. Attractiveness was determined by ratings of *S*s as to how much they liked or disliked doing each task.

[2] Attractiveness is defined in terms of preference. *S*s were asked to rank the experimental tasks in the order in which they liked to do them.

preference for the successful one. Rosenzweig (11) in an experiment involving 37 children from about five to 14 years of age found that when two simple jig-saw puzzles were done as a competitive task, one puzzle resulting in success, the other in failure, 32 of the 37 *Ss* reported that they liked better the puzzle on which success was met. Nowlis, working with undergraduate men, each of whom performed two puzzles and who were then asked which type of puzzle they preferred doing, observed that: "Success stimulation produces significantly more judgments of preference for a task than does Failure stimulation" (7, p. 324).

However, if the hypothesis stated above is correct, it seems necessary to account for the apparent contradiction found in observations and studies in which *Ss* provided with an opportunity to repeat the same tasks tended to repeat their failures. Successful performance appears not to have increased but to have decreased the attractiveness of the activities in that there was little tendency to repeat the successfully performed activities. The problem here appears to be a methodological one for these observations (1, 6, 11) were largely drawn from situations in which preference for performing activities was measured by noting whether the individual tended to repeat immediately the very same task he had just performed. As is stated by Rice in a recent disussion of the law of effect:

> The Law (of Effect) cannot be upheld if it means that success or satisfaction necessarily enhances the tendency for either or both (a) the specific response sequence to be repeated, and/or (b) the particular, or even the specific, goal object to be chosen again. Repetition of *both* these elements occurs chiefly in very young children, in mental defectives and in compulsive neurotics. With normal adults, both are repeated mainly in the case of routine habits or recurrent appetitive drives. . . . When normal adults repeat one of these two elements—the response as a whole or the choice of the particular goal object—after success, it is usually with a variation in the other. . . . Or, we may vary both the response and the goal, so far as the details go, while repeating with variations the same *kind* of response and selecting the same *kind* of goal that brought us success in the past (10, pp. 309–310).

As Rice further points out, these situations have tended to involve problem-solving in which the opportunity to vary response or goal is limited. Furthermore, measurement of attractiveness solely in terms of immediate repetition does not adequately consider all the factors in operation. A decrease in desire to repeat a performance may mean only temporary satiation and conceal a marked increase in the desire to perform that task or a similar task at some later time.

One other factor demands consideration. In experiments involving success and failure, success and failure are usually defined in terms of experimental operation. This by no means guarantees that the *Ss* will subsequently have genuine experiences of success and failure, and it has been suggested (2, 9) that ambiguity in experimental results may have its origin in uncontrolled variation in the *Ss'* actual experience. If success and failure are to be defined in terms of experimental operation it is clearly

evident that careful selection and application of the experimental techniques are necessary to strengthen the possibility of inducing genuine experiences of success and failure. Furthermore, some sort of experimental check or verification of the *S*s' experience seems desirable. All these factors have been given consideration in the design of the present experiment.

2. *Expectation.*—It is predicted that the general attractiveness of an activity will rise with the expectation of success and fall with the expectation of failure. The importance of expectation in the tendency to resume an interrupted task has been noted by Ovsiankina (8), Nowlis (7) and others. Tendencies toward resumption, which are normally strong, are reduced if the individual foresees certain failure. Grosslight and Child noted that failure interspersed with success in the operation of a mechanical apparatus led to greater persistence of unrewarded effort than when no previous failure was experienced. They tentatively suggest that this greater persistence may result from a greater expectation of success. ". . . *S*s have a uniform expectation that when a machine has once failed to work, persistence on their part will not be effective; . . . this prior expectation may be broken down by the experience of success after failure" (5, p. 387). The importance of expectation in the preference rating of interrupted activities was suggested to Cartwright in post-experimental interviews: "In conclusion, there is some indication, although the evidence is not statistically conclusive, that anticipation of failure more unequivocally reduces attractiveness than the actual experience of failure. This fact suggests that a person's view of his future is the primarily decisive factor in producing changes in attractiveness" (p. 13). Cartwright's results also indicate that tasks of the same category as the one completed, although not similar enough to insure certainty of success, become more attractive because of their greater potentiality of success.

In simultaneous consideration of both experience and expectation another hypothesis is advanced. As both experience and expectation lead to changes in attractiveness, it would appear to follow that these changes should be greater when experience and expectation are in the same direction, both of success or both of failure, than when they are in opposite directions, one of success and one of failure. When experience and expectation are in opposite directions the situation is more complex than when they are in the same direction. It has been predicted that the individual will find an activity more attractive if he has experienced success in it and find it less attractive if he has experienced failure. However, when circumstances strongly suggest that, despite the experience of success or failure in an activity, the opposite is to be expected in the future, the attractiveness of the activity need not be determined solely by past experience but may well be influenced by potential or expected success and failure.

Although experience and expectation have been discussed separately, this is not to imply that they are to be considered as two separate entities. Expectation of future success or failure appears to be determined by any number of factors, prominent among which is the past experience of success or failure. But experienced success and failure do not inevitably lead to expectations of success and failure respectively. Hence, there arises the

necessity of treating them separately to determine experimentally their separate and conjoint effects upon the attractiveness of activities.

3. *Strength of Need.*—When there is but little need to achieve success and to avoid failure in the performance of an activity it is predicted that changes in the attractiveness of that activity will follow the relatively simple pattern presented by experience and expectation; attractiveness increasing with the experience and expectation of success and decreasing with the experience and expectation of failure. When, however, this need is strong it is predicted that the attractiveness of an activity will not only increase with success and decrease with failure, but that these changes will be greater in both directions when the need is strong than when it is relatively weak.

It seems apparent that the circumstances surrounding the performance of a given activity may be such as to cause an individual to become 'involved' in the activity in such a way that his success or failure in it may be of considerable concern to him. In other circumstances when this 'involvement' is relatively weak, his success or failure in the same activity may be of only minor significance to him. These two situations are commonly described by the statement that the individual is more strongly motivated in the first situation than in the second. However, as an individual may have varying motives in the performance of any given activity, it would be more accurate to say that his motivation is greater in that he has a stronger need to achieve success and to avoid failure in the first situation than in the second.

In general, previous experimental work involving success and failure has attempted to establish such a need by offering incentives or by employing various techniques designed to involve the 'ego.' It is questionable whether all these techniques have succeeded in equal measure. Sanford (15), for example, suggests that the discrepancies between his and Rosenzweig's (14) results may in part be due to different degrees of ego-involvement between their two groups of *S*s. It seems reasonable to suppose that changes in the attractiveness of activities, coincident upon success and failure, may be a function of the presence or absence of this need. Consequently, direct experimental variation of this need merits study.

Experimental Design and Procedure

Nine different tasks were employed in the experiment. These tasks, either borrowed from or suggested by tasks appearing in currently used aptitude tests, were as follows: verbal arithmetic, codes, cube counting, sentence completion, paper formboard, number checking, sentence matching, symbol transcription, algebraic reasoning. The criteria used for the selection of these tasks and the details of a preliminary experiment made to standardize their difficulty can be found in the appendix. The series of nine tasks was presented to each *S*, who, after being informed that the experiment would require her to do some [3] of these tasks, was asked to rank them in the order in which she liked to do them.[4] The task given middle ranking (ranked fifth in a series of nine) by the *S* was selected by the *E* as the first one to be performed. To control the strength of need, half of the *S*s were told the experiment studied the relationship between interest

[3] Although the experimental design demanded only the performance of one task by each *S*, in most cases a second task of the *S*'s choice was performed and was so controlled by the *E* that *S*s left with feelings of success and general good-will toward the experiment.

[4] This wording was employed consistently to avoid the implication of immediate repetition of the same task.

and aptitude and that the selected task measured general intelligence; the other half were told the *E* sought information about the nature of the task and their reactions to it but that the actual quality of their performance was a matter of indifference (cf. 11). As all the tasks consisted of the same number of units or sub-tasks, all *S*s could be and were interrupted after seven of the sub-tasks were completed. Through comments of the *E* and the employment of purported norms, half of the *S*s were given reason to experience success, and half to experience failure. Again through comments of the *E* and interpretation of symbols which were printed on the back of each card and which represented two degrees of difficulty, half of the *S*s had reason to believe that chance had operated so as to give them an unusually difficult task, and half that chance had operated so as to give them an unusually easy task. Their expectations of success or failure in future performances of the same type of task might well vary accordingly. *S*s were then asked to make a second ranking of the tasks.

With two conditions of strength of need, two of experience, and two of expectation, there were possible eight combinations of experimental variables, and the experiment was planned according to the principles of factorial design. Ten *S*s were randomly assigned to each of the eight different conditions, making a total of 80 *S*s employed. The *S*s were all young women registered as students at Beaver College, Jenkintown, Pa. Volunteers were solicited by verbal appeal [5] in elementary psychology and business administration classes, and, with but a few exceptions, their backgrounds in psychology were limited to the first year course in general psychology. The content of this course up to the time of the experiment did not include material on intelligence testing, levels of aspiration, or other information which the *E* felt would bias the results of the experiment. Intercommunication between *S*s was discouraged by stating frankly during the appeal that results would be influenced and, hence, rendered worthless if the *S*s had fore-knowledge of the experiment. It was requested that no one volunteer who felt she could not refrain from talking about it. After each *S* had completed the experiment she was again reminded of the importance of not talking about it. From the many spontaneous remarks [6] there is reason to believe that cooperation was generally excellent. To avoid the possibility of traumatic effect, *S*s who were placed in the condition of experienced and expected failure and any others who appeared disturbed by the outcome of the experiment were taken into confidence as to the true purpose of the experiment before they left the experimental room.

The measures of attractiveness were the two preferential rankings of the tasks made before and after performance of the critical task. The changes in the attractiveness of the critical task were determined from the second rankings. As the critical task was the one given the rank of five in the first ranking, a rank smaller than five in the second ranking would indicate an increase in attractiveness, larger than five, a decrease in attractiveness.

Two questionnaires were used, one of which was designed to check the successful induction of the three experimental variables; the other to ask general questions justifying to the *S*s the instructions used to establish the need condition.

The experimental procedure in detail follows. The verbal instructions were memorized by the *E* but a copy was always kept in sight for surreptitious reference.

Introduction

When the *S* entered she was greeted informally and seated at a long table on which lay the nine tasks.[7]

---

[5] This was supplemented by a display of attractive hand-painted plaques which were to be presented them in appreciation of their cooperation. *S*s received these plaques as they were leaving the experimental room. The gift was in no way connected with the quality of their performance.

[6] Typical examples of these comments are as follows:

S44: "It's funny how the girls don't talk about it. No one will say anything about it. You know they usually do talk when they're asked not to."

S53: "You know it has been a whole week and no one has said anything about it to me. I thought you'd like to know that. My girl friend lives across the hall from me but she refuses to say anything about it at all."

S64: "I asked some of them how it was, but they only smiled and said, 'Don't worry about it.'"

[7] Each task consisted of a series of problems, each problem printed on a separate card. For convenience the cards were kept in shallow, open boxes which were identified by the letters from

"The experiment will require your doing some of these tasks. Sometimes I will select the task, other times your choice will determine the task. Before we begin then, it is necessary to know what activities you like to do. I think you are probably familiar with all these tasks, but just to make sure, let me explain them to you."

Description of Tasks

Descriptions of the tasks were similar to those found in current aptitude tests.

First Ranking

"I am going to ask you to list these tasks in the order in which you like to do them. Make the one you like to do best number one, the one you like to do second best, number two, and so on down, with number nine being the one you like to do least. Do you understand? You can designate the tasks by their letters [on box]."

S made first ranking which was immediately removed and placed face down on the table.

Establishment of Need

Little Need: "First, you really ought to know something about the purpose of this procedure. In almost every psychological experiment the *S* is asked to do some task and the *E* is very careful to choose a task that will help him get the information he wants to get. The tasks you see here are all possibilities in an experiment I am going to do very shortly. To make sure that the results have a good chance of turning out the way I want them to, I must get more information about the nature of each task and how people react to it, before I select the task I am to use. Now, as you were willing to help out in this work you will be asked to do some of these tasks. You need not be concerned about how well or how poorly you do. This isn't a test and I don't care how well or how poorly you do. What I want you to do is to carefully observe your reactions to the tasks themselves, so that you can answer some general questions I'm going to ask you afterward about them. O.K.? All right. I'll select the first task." [*E* shook a box filled with folded slips of paper, drew a slip from the box, unfolded it, and announced] "The first task is task ..... [letter designation]." [The selected task was always the task to which the *S* had given the middle ranking. Its identity was determined by casual observation as the *S* completed her first ranking.]

Strong Need: "First you really ought to know something about the purpose of this procedure. This experiment has been planned as a study on the relationship between ability to do a task and interest in it. I want to find out how well people do in tasks in which they are interested as compared to how well they do in tasks in which they are not interested. Each of the tasks on this table measures a different ability. How well or how poorly you do will be measured by your performance and this will indicate how much of that ability you possess. O.K.? All right. I'll select the first task." [Procedure of selecting task same as above.] "Oh!" [pause] "This ought to be a challenge to you. The first task will be task ..... Task ..... measures general intelligence. Tests like it are sometimes used in entrance tests given to determine ability to do college work. Do the very best you can on it."

Beginning of Task

[*E* picked up all the cards in the designated box and started to shuffle them.] "I'm going to give you 10 of these cards." [pause] "I'm going to stop you after a while so I'll need a stop watch." [*E* with cards in hand stepped to side of room and opened a large closet door which opened at an angle so as to hide the contents of the closet and most of *E*'s movements. A stop watch was picked up and a set of 10 cards of the same type but of prearranged difficulty were slipped on top of the pack still in the *E*'s hands. *E* returned to table, counted off 10 cards from the top of the pack and gave them to *S* with an answer sheet.] "Each card has a different number. As you do each one always write down the number of the card and put your answer beside it. All right? Start."

Ending of Task

After seven of the 10 problems were completed all *S*s were stopped.

Strong Need: "Stop. Read me your answers." [As *S* read the answers *E* checked them

---

A to I. The top card in each box was marked 'sample' and was used to complement the verbal description of the task. The boxes were arranged in two horizontal rows (five in the first), about a foot from the *S* and about two in. apart. The position of all tasks remained constant throughout the experiment.

against answer book but did not indicate for each item whether *S* was right or wrong.] "That's O.K. Let's see how well you did." [*E* picked up booklet of norms.] [8]

Little Need: "Stop. Now I want you to answer some questions on your reaction to this task. Where is that sheet? I had it just a minute ago." [*E* searched through a series of folders on top of the experimental table and on an adjoining table.] "This isn't it either." [*E* came upon the booklet of norms and laughed shortly.] "I didn't know I had this with me. One of my friends at school used these same tasks as a test and gave me a copy of his results." [*E* was about to replace the booklet when she suddenly reversed movements and said] "Let's see how well you did, just for the fun of it." [*E* checked name of task and time on stop watch and then looked in booklet.]

Induction of Experience and Expectation of Success and Failure

As there were four possible combinations of these variables the instructions vary accordingly.

Experience Success; Expect Success

"Fifty percent of all college students tested solved four in this time. How many did you do? [Pause for reply] Seven! [*E* checked *S*'s paper.] That's right. That's really good. That puts you up near the top of the group." [Norms were briefly shown to *S* with *E* pointing to the score of seven which appeared near the top.]

"Let me see the cards for a moment." [*E* turned cards face down and spread them out.] "My word! Look at this! I wanted to see how difficult the problems were that you got. This symbol [9] indicates that the problem is one of the easier ones, this symbol that it is one of the more difficult ones. You can see that you have had a *bad* break in that most of yours were difficult ones, many more than is usual, and despite that you have done a good job! If you would have had a different set you probably would have done even better than that. Nice going!"

Experience Success; Expect Failure

[Same as above except for] "You can see that actually you have had a *lucky* break in that most of yours were easy ones, many more than is usual. Now the fact that you had easy ones doesn't discredit your success, that stands! It does mean that if you would have had a different set you probably wouldn't have done nearly so well. You certainly had a lucky break!"

Experience Failure; Expect Success

"Fifty percent of all college students tested solved all 10 in this time. How many did you do? [Pause for reply] Only seven!" [*E* checked *S*'s paper.] "That's right. Ummm! That's not so good. That puts you down, near the bottom of the group." [Norms were briefly shown to *S* with *E* pointing to the score of seven which appeared near the bottom.]

"Let me see the cards for a moment. [*E* turned cards over and spread them out.] My word! Look at this! I wanted to see how difficult the problems were that you got. This symbol indicates that the problem is one of the easier ones, this symbol that it is one of the more difficult ones. You can see that you have had a *bad* break in that most of yours have been difficult ones, many more than is usual. No wonder you did so poorly. If you would have had a different set you probably would have done much better. That really was a bad break!"

Experience Failure; Expect Failure

[Same as above except for] "You can see that you have had a *lucky* break in that most of yours have been easy ones, many more than is usual. You really have done a poor job. If you would have had a different set you probably would have done even poorer. I don't know!"

Transition to Second Ranking

"Well, we really shouldn't have taken the time out to talk about that but it's always interesting to know how well you've done. Let's go ahead."

---

[8] Two sets of norms were used to meet the varying requirements for the induction of the experiences of success and of failure.

[9] All cards had either an asterisk or a small square printed on the back. Eight of the 10 cards had an asterisk, two a square. If the task was to be interpreted as easy the asterisk was said to indicate an easy problem. In interpreting the task as difficult the asterisk was said to indicate a difficult problem.

Second Ranking

"Sometimes when you have had an opportunity to actually do tasks you change your mind about how much you like to do them. I want you to do now as you did before, list the tasks in the order in which you like to do them. Make the one you like to do best number one, the one you like to do second best, number two and so on down. I don't care whether or not you are consistent. Just so your rankings indicate exactly how you feel about them now. You are to include in your ranking this task you've just worked on. I'll put away this one set of problems you've just done."

Questionnaires

"I need some other information on how interesting the task is which will involve your answering this series of questions. Read the instructions and answer the questions as carefully as you can." [*S* was given questionnaire No. 1 which included the following questions to be answered on a seven-point rating scale: 1. How interesting do you feel this task is to *people in general?* (Very interesting, interesting, somewhat interesting, neither interesting nor uninteresting, somewhat uninteresting, uninteresting, very uninteresting) 2. Is this task as interesting to *you* as you anticipated? (Much more interesting, more interesting, somewhat more interesting, as expected, somewhat less interesting, less interesting, much less interesting) 3. What comments, suggestions or criticisms can you offer with respect to this task, or to the experimental situation itself? In brief, what is your reaction to it?]

[After *S* completed questionnaire No. 1] "Here is another series of questions to answer." [*S* was given questionnaire No. 2 which included the following questions: 1. How well do you feel you have done on this task? (Very superior, superior, somewhat above average, average, somewhat below average, inferior, very inferior) 2. It was originally planned to have you repeat each task you undertook. How well would you expect to do if the same type of task were repeated? (Very superior, superior, somewhat above average, average, somewhat below average, inferior, very inferior) 3. To what extent do you feel satisfied with your performance of this task? (Very satisfied, satisfied, rather satisfied, neither satisfied nor dissatisfied, rather dissatisfied, dissatisfied, very dissatisfied.)]

Conclusion

S was then asked to pick a task she would like to do and a modified form of the experimental procedure was roughly followed, including an experience of success. In some cases where time did not permit a second task in the alotted half-hour this was explained to the *S* and the experiment was brought to a conclusion at this point.

Results and Discussion

I. *Analysis of Attractiveness Rankings of Critical Task*

The critical task for each *S* was the task given the mid-rank of five in the first rankings of attractiveness. Changes in attractiveness were determined by noting the position of the critical task in the second rankings made after the performance of the task. A value smaller than five indicates an increase in attractiveness, a value larger than five a decrease in attractiveness. Table I gives the mean second rank of the critical task for the 10 *S*s in each of the eight conditions. It is to be observed that there is, overall, a slight tendency for the second rankings to indicate increased attractiveness of the critical task; the mean of all second rankings being 4.8. However, a *t*-test to determine whether the second rankings were significantly different from the first, i.e., significantly smaller than five, yields a *P*-value of $>.05$ which does not indicate statistical reliability.

In Table I it is possible to compare groups of Ss who were assigned identical conditions of need and expectation but different conditions of *experience*. In each of the four comparisons which can thus be made, the mean ranks show that the critical task was considered more attractive with an experience of success than with an experience of failure (compare 3.7 and 4.9, 4.0 and 4.6, 5.0 and 5.2, 4.6 and 6.5) Likewise there are possible four comparisons between groups of Ss assigned identical conditions of need and experience but different conditions of *expectation*. Again in each comparison the mean ranks show that the task was considered more attractive with an expectation of success than with an expectation of failure (compare 3.7 and 5.0, 4.0 and 4.6, 4.9 and 5.2, 4.6 and 6.5). In comparisons between groups of Ss with identical conditions of experience and expectation

TABLE I

MEAN RANK OF ATTRACTIVENESS OF THE CRITICAL TASK ON NINE-POINT SCALE AFTER PERFORMANCE UNDER EACH EXPERIMENTAL CONDITION

	Experience Success		Experience Failure	
	Expect Success	Expect Failure	Expect Success	Expect Failure
Strong Need	3.7	5.0	4.9	5.2
Little Need	4.0	4.6	4.6	6.5

but different conditions of *strength of need*, there appears to be no consistent direction of difference in the attractiveness of the activity under strong and little need conditions (compare 3.7 and 4.0, 5.0 and 4.6, 4.9 and 4.6, 5.2 and 6.5).

To determine the statistical reliability of these differences resulting from the experimental variables of experience, expectation and need, an analysis of variance was made. The probabilities that these differences would occur in random sampling appear in Table II. The table indicates clearly that the second rankings of the critical task following the condition of experienced success are significantly higher than those following the condition of experienced failure ($P < .01$). Likewise, the second rankings under the condition of expected success are significantly higher than those under the condition of expected failure ($P < .01$). The difference between the conditions of need and all interactions, both double and triple, are statistically without significance.

Changes in attractiveness can also be presented in terms of the number of Ss whose second preference ranking of the critical task show increased, decreased, or unchanged attractiveness. If a chi-square test is applied to a 3×2 table subclassified as to the frequency of

TABLE II

ANALYSIS OF VARIANCE OF ATTRACTIVENESS RANKINGS OF CRITICAL TASK AFTER ITS PERFORMANCE

Experimental Variables	df	Sum of Squares	Variance	*F*	*P*
Experience	1	20.00	20.00	8.55	<.01
Expectation	1	22.05	22.05	9.42	<.01
Need	1	.80	.80	2.92*	>.05
Exper × Expec	1	.20	.20	11.70*	>.05
Exper × Need	1	1.25	1.25	1.87*	>.05
Expec × Need	1	.80	.80	2.92*	>.05
Exper × Expec × Need	1	6.05	6.05	2.59	>.05
Error	72	168.40	2.34		
Sum	79	219.55			

* With 1 and 80 degrees of freedom an *F* of 3.96 has a *P*-value of .05; an *F* of 6.96 has a *P*-value of .01. When the error variance exceeds the experimental variance and with 75 and 1 degrees of freedom an *F* of 253 has a *P*-value of .05; an *F* of 6,323 has a *P*-value of .01.

rankings showing increased, decreased or unchanged attractiveness and as to experience of success or failure, the resultant *P*-value is less than .08. A similar test between changes of attractiveness and expectation yields a *P*-value of less than .05. If the same data are presented in 2 × 2 tables with the decreased and unchanged attractiveness combined into one class, then the chi-square test applied to changes in attractiveness and experience give a *P*-value of less than .02 and applied to changes in attractiveness and expectation give a *P*-value of less than .08. All four values either have, or closely approach statistical reliability, with expectation possessing greater statistical significance under the three degree classification of changes in attractiveness, and experience under the two degree classification. All these differences are in the same direction as those obtained from the mean rankings. With experience or expectation of success more rankings of the critical task show increased attractiveness than either unchanged or decreased attractiveness. With the experience or expectation of failure more rankings show decreased attractiveness than either unchanged or increased attractiveness.

Thus, whether one considers the means of the second rankings of attractiveness or the numbers of Ss whose rankings of attractiveness show changes in a given direction, the results of the present experiment consistently show that with the conditions of experienced and expected success there is an increase in the attractiveness of the critical task and with the conditions of experienced and expected failure there is a decrease in the attractiveness of the critical task. As the data are decisive with respect to the stated experimental hypotheses, they also tend to confirm and clarify Cartwright's hypotheses as to the effect of experience and expectation on his own re-

sults. Attractiveness of a task is determined not only by present success or failure but by expectations of future success or failure as well. Consequently, it is quite possible under certain conditions of expectation for the attractiveness of the task to rise even when the *S* has experienced failure and for the attractiveness of the task to fall even when the *S* has experienced success. In the present experiment itself, the attractiveness of the task rose for eight of the 20 *S*s under the conditions of experienced failure, expected success, and fell for eight of the 20 *S*s under the conditions of experienced success, expected failure.

It was predicted that the changes in attractiveness would be *greater* when the experience and expectation were in the same direction, both of success or both of failure, than when they were in opposite directions, one of success and the other of failure. When the averages in Table I are combined without respect to strength of need, the average ranking of attractiveness of the critical task for experienced and expected success is 3.85, for experienced and expected failure, 5.9. A difference of 2.05 separates the two averages. A *t*-test applied to the data yields a *P*-value of less than .01 and indicates that the difference between the two sets of data is statistically reliable. The average ranking of attractiveness for experienced success and expected failure is 4.8, for experienced failure and expected success, 4.75. The difference of .05 which separates them is obviously not significant. The difference between the differences of 2.05 and .05 is in itself significant with a *P*-value of $<.01$. [The hypothesis that experience and expectation in the same direction would result in greater changes in attractiveness then would experience and expectation in opposite directions appears confirmed by the above data.

The *directions* of change in attractiveness where experience and expectation are in the same direction are as anticipated; attractiveness increases with the experience and expectation of success and decreases with the experience and expectation of failure. When experience and expectation are in opposite directions, there is no evidence in the rankings to indicate the dominance of either expectation or experience in determining the direction of the changes in attractiveness.] Of the average values which were quoted above, those obtained under the conditions of experienced failure and expected success do show a slight average rise in attractiveness (4.75) but the values obtained under the conditions of experienced success and expected failure also show a slight average rise in attractiveness (4.8). As there is considerable variability in both these groups which is not revealed by their averages, it is possible that the absence of a predominant direction of change when experience and expectation are in contrast may be due to actual variation among the *S*s as to the

relative strength of the experience and expectation, some *S*s responding largely on the basis of their experience, some on the basis of their expectation.

The prediction that changes in attractiveness would be greater in either direction under the condition of strong need than under the condition of little need is not confirmed. As mean changes in attractiveness are determined by observing the distance of the mean second rankings from the mid rank of five, it is apparent that the greatest increase in attractiveness occurs under the conditions of experienced and expected success with a strong need, the mean rank of 3.7 indicating an average increase of 1.3. The remaining values under the condition of strong need show little or no change in attractiveness, 5.0, 4.9, 5.2. The greatest average decrease in attractiveness occurs with experienced and expected failure with little need, a mean rank of 6.5 indicating an average decrease of 1.5. The remaining values under little need show increases in attractiveness, 4.0, 4.6, 4.6. The difference between the highest and lowest average changes under strong need is 1.5, under little need 2.5. When separate analyses of variance are made on the data within the strong need and little need conditions, the *P*-values are statistically significant for experience and expectation within little need, but not significant within the condition of strong need. However, it is to be recalled that no significant interaction between experience and need, nor between expectation and need, appears in the main analysis of variance. There is no significant interaction between experience and expectation under either the strong or little need conditions. The present data are not adequate to account satisfactorily for these results and for their failure to support the hypothesis that changes would be greater in either direction under the strong need condition than under the little need condition.

II. *Analysis of General and Personal Interest Ratings of Critical Task*

As a justification to the *S*s for the instructions employed to establish the need conditions, a questionnaire was introduced after the performance of the critical task which required each *S* to rate the interest value of the task both for people in general and for herself personally. With a scale range of seven points, a value smaller than four denotes a judgment of interesting, larger than four, of uninteresting. Table III gives the mean ratings of general and personal interest in the task for all eight conditions. As indicated in Table III there is an overall tendency to rate the task as interesting, no single average being a value larger than four. It is to be observed that the critical task is consistently rated as having greater general

TABLE III

MEAN RATING ON SEVEN-POINT SCALE OF GENERAL AND PERSONAL INTEREST IN CRITICAL TASK AFTER PERFORMANCE UNDER EACH EXPERIMENTAL CONDITION

		Experience Success		Experience Failure	
		Expect Success	Expect Failure	Expect Success	Expect Failure
Rating of General Interest	Strong Need	2.4	2.5	2.4	2.5
	Little Need	3.0	2.7	2.8	3.1
Rating of Personal Interest	Strong Need	1.9	2.7	3.4	3.4
	Little Need	2.5	2.7	3.1	3.9

interest by *Ss* under the conditions of strong need than by *Ss* under the condition of little need (compare 2.4, 2.5, 2.4, and 2.5 with 3.0, 2.7, 2.8, and 3.1, respectively). This distinction on the basis of strength of need does not appear in ratings of personal interest. Here, rather, the interest appears to vary directly both with experience and expectation. The task is rated as having greater personal interest when success is experienced than when failure is experienced (compare 1.9, 2.7, 2.5, and 2.7 with 3.4, 3.4, 3.1, and 3.9, respectively). The task is, likewise, rated as having greater personal interest when success is expected than when failure is expected (compare 1.9, 2.5, 3.4, and 3.1 with 2.7, 2.7, 3.4, and 3.9, respectively). The results of the analysis, to be found in Table IV (columns 2 and 3), demonstrate the statistical reliability of the observed differences in the ratings of general interest between need conditions ($P < .05$) and of personal interest within experience and within expectation ($P < .01$ and $= .05$). There are no significant double or triple interactions.

Below the ratings of general and personal interest, space was

TABLE IV

PROBABILITIES FROM ANALYSIS OF VARIANCE OF INTEREST AND VARIABLE INDUCTION RATINGS AFTER PERFORMANCE OF CRITICAL TASK

Experimental Variables	Ratings of		Ratings of		
	General Interest	Personal Interest	Experience	Expectation	Satisfaction
Experience	>.05	<.01	<.01	<.01	<.01
Expectation	>.05	=.05	>.05	>.05	>.05
Need	<.05	>.05	>.05	>.05	>.05
Exper × Expec	>.05	>.05	>.05	>.05	>.05
Exper × Need	>.05	>.05	>.05	>.05	<.05
Expec × Need	>.05	>.05	>.05	>.05	>.05
Exper × Expec × Need	>.05	>.05	<.01	<.01	<.05

provided on the questionnaire for comment by the S. The request for comment was broad enough in scope to allow the Ss to state freely their reactions to the task. These statements were typed on sheets of paper and identified only by numbers. Five judges, all graduate students in psychology, independently rated each response as favorable, neutral (or indeterminate), or unfavorable with respect to the critical task, the experimental situation, or S's estimate of his own performance or ability. A comparison of all five series of judgments showed 39 cases of complete agreement, 14 cases in which four judges agreed, 14 cases in which three judges agreed and three cases in which there was no majortiy of agreement. In addition there were 10 cases in which the Ss failed to comment. The three cases of disagreement were placed in the neutral or indeterminate classification. This gave 29 favorable, 13 neutral, and 28 unfavorable judgments in addition to the 10 cases of no comment. The averages of the second preference rankings of the critical tasks were computed for each group and they are as follows: favorable comment, average ranking 3.7; neutral or indeterminate, 4.7; unfavorable, 5.4; no comment, 5.8. There is a gradual decrease in attractiveness from the favorable comments to the unfavorable comments, with the lowest average ranking coming from those who declined to comment. The difference in ranking between favorable and unfavorable comments yields a C. R. of 4.14 and a *P*-value of less than .01 which signifies statistical reliability. A chi-square test of a 2×2 contingency table, subclassified as to favorability or unfavorability of comment and as to experience of success or failure, yields a *P*-value of less than .01. A similar test between favorability and expectation yields a *P*-value of less than .05. Both indicate the existence of significant differences in favorability of comments; the comments being more favorable with the experience and expectation of success and less favorable with the experience and expectation of failure. No such difference exists with respect to need.

It is of interest that the trends in the above results are in the same direction as those found in the second rankings of attractiveness; personal interest in the task is greater and comments more favorable when success is experienced or expected than when failure is experienced or expected.

III. *Analysis of Ratings of Experimental Variables*

As a partial check upon the effective induction of the experimental variables, a second questionnaire was introduced after performance of the experimental task which required each S to rate the direction and strength of her experience, expectation and satisfaction with re-

sults. With a rank of four indicating the mid-point of a seven-point scale, a value smaller than four represents an experience or expectation of success, or satisfaction with results, while a value larger than four represents an experience or expectation of failure, or dissatisfaction with results.

Table V gives the mean ratings of experience, expectation and satisfaction made by the *Ss* in each condition. The table discloses the fact that the experimental variable of experience appears to have influenced all three ratings. The experimental condition of experienced success resulted in ratings generally smaller than four which may be taken to indicate the actual experience of success (3.4, 4.0, 3.5, 3.7), the expectation of success (3.0, 3.4, 3.2, 3.2), and satisfaction

TABLE V

MEAN RATING ON A SEVEN-POINT SCALE OF EXPERIENCE, EXPECTATION AND SATISFACTION MADE AFTER PERFORMANCE UNDER EACH EXPERIMENTAL CONDITION

		Experience Success		Experience Failure	
		Expect Success	Expect Failure	Expect Success	Expect Failure
Rating of Experience	Strong Need	3.4	3.5	5.7	5.4
	Little Need	4.0	3.7	4.4	6.0
Rating of Expectation	Strong Need	3.0	3.2	4.8	4.2
	Little Need	3.4	3.2	3.5	4.7
Rating of Satisfaction	Strong Need	2.2	3.0	5.9	5.3
	Little Need	3.0	3.2	4.6	5.5

with results (2.2, 3.0, 3.0, 3.2). The experimental condition of experienced failure resulted in ratings generally larger than four which may be taken to indicate the experience of failure (5.7, 4.4, 5.4, 6.0), expectation of failure (4.8, 3.5, 4.2, 4.7) and dissatisfaction with results (5.9, 4.6, 5.3, 5.5). Neither the experimental variables of expectation nor of need alone appear to have had a systematic effect upon any of the three ratings.

In Table IV (columns 4, 5, and 6) it is to be observed that the differences within the ratings of experience, expectation and satisfaction, produced by the experimental variables of experience, possess statistical reliability ($P < .01$ in each case). These differences due to experience strengthen belief in the successful operation of the variable of experience. The absence of a similar effect of the experimental variable of expectation upon rated expectation was not anticipated since the variable of expectation had produced statistically reliable

differences in the second rankings of attractiveness of the critical task, in the ratings of personal interest, and in the favorability of the written comments. The present experimental data are not adequate to test most of the explanations which suggest themselves here. However, an analysis of covariance between the ratings of experience and expectation, holding experience constant, completely cancels all significant differences within the ratings of expectation. Whatever factors operated in the ratings of experience appear to have been carried over to the ratings of expectation. An analysis of covariance between the ratings of experience and expectation, holding experience constant, cancels all significant differences within the ratings of satisfaction except that induced by the experimental variable of experience itself.

It should be noted that in all three ratings there are significant triple interactions of experience, expectation and need. All three interactions follow a similar pattern. When a *strong* need is present the ratings are high for experienced success and considerably lower for experienced failure. The expectation has only little effect, but its effect is *inverse* for experiences of success and failure, in that when success is experienced the lower ratings accompany an expectation of failure, but when failure is experienced the lower ratings accompany an expectation of success. When *little* need is present, again the ratings are higher for experienced success than for experienced failure but reach neither the height nor depth occuring with strong need. Here expectation has a somewhat stronger effect, but its effect is *inverse* for experience of success and failure and *inverse to the results of strong need.* When success is experienced the lower rating accompanies an expectation of success. That these three triple interactions are not independent is again suggested by the analysis of covariance between ratings of experience and expectation, and between ratings of experience and satisfaction with results, for the triple interaction fails to appear when experience is held constant.

Summary and Conclusions

1. When a task of relatively neutral attractiveness was performed, changes in attractiveness were dependent upon *experience* of success and failure; experienced success, in general, was accompanied by a rise in attractiveness and experienced failure by a fall in attractiveness.

2. When a task of relatively neutral attractiveness was performed, changes in attractiveness were dependent upon *expectation* of success and failure; expected success, in general, was accompanied by a rise in attractiveness and expected failure by a fall in attractiveness.

3. Average changes in the attractiveness of the task were significantly greater when the expectation and experience were in the same direction, both of success or both of failure, than when experience and expectation were in opposite directions, one of success and the other of failure.

4. When changes in attractiveness were considered separately for strong and little need, experience and expectation both produced significant variation in the mean preference rankings under little need but not under strong need.

5. Ratings of personal interest and favorability of general comments were dependent upon experience and expectation of success and failure; personal interest was greater and general comments more favorable when success was experienced or expected than when failure was experienced or expected.

6. Ratings of experienced success or failure, expected success or failure, and satisfaction with results, appeared to be determined largely by the experimental condition of experience.

7. In general, it was shown, as was suggested by Cartwright, that the attractiveness of an activity is determined not only by past experience of success or failure but by expectation of future success and failure as well.

Appendix

Selection of Tasks

The tasks selected were ones which met the following criteria:

1. Sufficiently interesting and challenging to be of interest to *S*s of college age and ability.

2. Capable of interpretation either as a test of intelligence or other ability or as a casual pastime activity.

3. Capable of presentation as a task made up of many units or sub-tasks rather than a single unified task. This last feature has these advantages.

(a) Allows interruption during the task, which is superior to completion in facilitating the induction of feelings of failure, and yet largely removes the danger of increased attractiveness attributable not to the experimental variables but to feelings of incompleteness.

(b) Allows employment of a large number of sub-tasks (30) for each type task from which is drawn a set (10) to be solved by the *S*. The remainder would allow *S* to state a preference for repeating the same type of activity without involving the repetition of the identical task.

(c) Allows the employment of a technique leading to the apparent operation of chance in determining the difficulty of the *S*'s set of sub-tasks.

4. Capable of uniform presentation; each sub-task on a 3 × 5 in. card.

5. Allows an accurate check on *S*'s progress for, as all are pencil and paper tasks, *S* writes the answer to each sub-task on a record sheet. This allows interruption of all *S*s at the same point in the task.

A preliminary experiment was performed for the purpose of standardizing the tasks, to enable the selection of tasks of relatively equal difficulty. From the 30 sub-tasks which were selected or constructed for each of the nine tasks, 20 within the desired range of difficulty were chosen. Twenty-two girls from the University of Pennsylvania, who volunteered to serve as *S*s, were each required to solve half of these problems so that each sub-task was performed eleven times in all. The measures taken included rankings of preference before and after the performance of all tasks, ranking of difficulty after performance, time score for each sub-task and correctness of answers. The correlation of the first and second rankings of preference made before and after performance was +.90. The rankings of difficulty correlated +.83 with the

first rankings of preference and +.92 with the second rankings of preference. The correlation of the median time with experienced difficulty was −.18. In selecting the final set of 10 problems the ratings of difficulty and correctness of answers were considered. Where a task was considered easy the more difficult problems were selected, if considered difficult, the easier ones. Items failed by large numbers of *S*s were avoided. Tasks selected as 'samples' were tasks of medium difficulty.

(Manuscript received July 8, 1947)

References

1. Allport, G. W The ego in contemporary psychology. *Psychol. Rev.*, 1943, **50**, 451–478.
2. Cartwright, D. The effect of interruption, completion, and failure upon the attractiveness of activities. *J. exp. Psychol.*, 1942, **31**, 1–16.
3. Festinger, L. A theoretical interpretation of shifts in level of aspiration. *Psychol. Rev.*, 1942, **49**, 235–250.
4. Frank, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 119–128.
5. Grosslight, J. H., & Child, I. L. Persistence as a function of previous experience of failure followed by success. *Amer. J. Psychol.*, 1947, **60**, 378–387.
6. Hoppe, F. Erfolg und Misserfolg. *Psychol. Forsch.*, 1931, **14**, 1–62.
7. Nowlis, H. H. The influence of success and failure on the resumption of an interrupted task. *J. exp. Psychol.*, 1941, **28**, 304–325.
8. Ovsiankina, M. Die Wiederaufnahme unterbrochener Handlungen. *Psychol. Forsch.*, 1928, **11**, 302–379.
9. Prentice, W. C. H. The interruption of tasks. *Psychol. Rev.*, 1944, **51**, 329–340.
10. Rice, P. B. The ego and the law of effect. *Psychol. Rev.*, 1946, **53**, 307–320.
11. Rosenzweig, S. Preferences in the repetition of successful and unsuccessful activities as a function of age and personality. *J. genet. Psychol.*, 1933, **42**, 423–441.
12. Rosenzweig, S. The recall of finished and unfinished tasks as affected by the purpose with which they were performed. (Abstract) *Psychol. Bull.*, 1933, **30**, 698.
13. Rosenzweig, S. An experimental study of 'repression' with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64–74.
14. Rosenzweig, S., & Mason, G. An experimental study of memory in relation to the theory of repression. *Brit. J. Psychol.*, 1934, **24**, 247–265.
15. Sanford, R. N. Age as a factor in the recall of interrupted tasks. *Psychol. Rev.*, 1946, **53**, 234–240.

STIMULUS IDENTITY AS RELATED TO RESPONSE SPECIFICITY AND RESPONSE GENERALIZATION

BY DELOS D. WICKENS

Ohio State University

Introduction

In several studies by the present writer (**6, 7**) the question of how the conditioned response is affected by perceptual factors at the time of its elicitation was raised. In these studies an extensor movement of the finger was conditioned, and then the arm was turned to a position so that the opposite or flexor movement would have to be made if the finger was to be removed from the electrode.

The electrode itself was placed upon a floating platform, so that either an extensor or a flexor movement could be recorded. The conditioned stimulus was then sounded, without, of course, there being any presentations of the unconditioned stimulus in this new position. In all of these studies the general trend of the results was the same. In all instances the majority of the *Ss* responded by making flexor rather than the originally conditioned extensor movements; a few *Ss* failed to respond at all; and a very small number of *Ss* responded by making extensor movements. In other words, although response generalization was the rule, there were a few instances of response specificity.

At the termination of the experimental session the *Ss* were questioned concerning their thoughts at the time the hand was turned over and the conditioned stimulus was sounded. Explanations of behavior were not readily forthcoming, and most of the *Ss* simply said that their fingers acted in the way they did more or less of their own accord; there was no satisfactory way of discriminating between the *Ss* who made extensor movements and those who made flexor movements on the basis of verbal reports.

Thus, under presumably identical conditions, two opposing types of response occurred. For the majority of the *Ss* the conditioned response which did occur was modified by the new environmental situation. The response had become generalized and was nonspecific in nature. In a few instances the original form of the response persisted even though this form would be classified as nonadaptive in the new situation.

The occurrence of these contradictory responses in the same experimental situation indicated the need for further research which

would be directed toward analyzing out the factors of which respones generalization in conditioning is a function. The present study is an attempt to investigate the interaction of stimulus generalization and response generalization. Specifically, is the generalized response more or less likely to occur if a tonal stimulus in the test situation is the same as or different from the stimulus used in the training situation?

Problem

Is response generalization affected by the nature of the conditioned stimulus used?

Subjects and Apparatus

The *S*s in this experiment were 58 volunteers from the elementary psychology class at the University of Wisconsin.

The conditioning board consisted of a plank about three feet by one foot. This was fixed with adjustable restraining devices so that the *S*'s hand and arm could be fastened securely in position. Straps were placed over the *S*'s forearm and wrist and the back of his hand. A block at the back of the elbow prevented the *S* from making horizontal withdrawal movements.

One of the shock electrodes was strapped to the *S*'s wrist, while his middle finger rested on the other electrode. The shock circuit could be broken by removing the finger from the electrode. The tone, used as the conditioned stimulus, was delivered through earphones worn by the *S*.

Finger responses were recorded by means of electrical contacts. A narrow strip of metal, hinged at one end, rested on the top of the *S*'s finger. When the finger was lifted off the electrode, this metal strip came in contact with a metal spring above it, thus closing the circuit. Downward movements of the finger could also be recorded. The electrode on which the finger rested was attached to a platform which was supported by a spring. The platform was pierced by a bolt. When the platform was depressed, the bolt dipped into a mercury well, and the recording circuit was closed. Both the upward and downward response circuits could be adjusted so that movements of the finger as small as $\frac{1}{16}$ in. could be recorded. Lights on a control panel in the *E*'s room indicated whether an upward or a downward response occurred.

The shock stimulus was delivered by a high voltage power pack. This, coupled with the facts that one electrode was strapped on the *S* and that the other was small in area, served to maintain the shock at a fairly constant value from stimulation to stimulation.

The tone was delivered by means of a Maico Audiometer.

The time relations between the conditioned and the unconditioned stimuli were controlled by a timing disk. This was a disk driven by a constant speed motor. On it were lugs which moved into contact with brushes, and, when contact was made, relays were closed and the stimuli were applied. The conditioned stimulus occurred for 0.5 sec., and was followed, 0.4 sec. after its onset, by the unconditioned stimulus. The unconditioned stimulus had a duration of about 0.1 sec. At a constant interval of time before the occurrence of the conditioned stimulus, a clock was started. This clock was stopped by the *S*'s response. In this manner latent times in hundredths of a sec. could be recorded.

S and *E* were seated in adjacent, sound proofed rooms. A one-way mirror between *S* and *E* permitted *E* to observe *S* without himself being observed. Switches on a control panel made it possible to administer the conditioned and the unconditioned stimulus independently of each other, as well as permitting variation of the inter-stimulus period.

Procedure

In general, the procedure of the experiment consisted of conditioning the *S*s while the hand was strapped palm down, so that the extensor movement removed the finger from the electrode. Following this, the hand was turned over, palm up, so that a flexor movement was required to lift the finger from the electrode. While the hand was in this new position, the conditioned stimulus

alone was sounded 10 times. For 27 of the *Ss* the same tone was employed in the transfer test situation as was employed in the training situation. The tone was 50 decibels above the threshold with a frequency of 256, 512, or 1024. For 32 of the *Ss* the test tone differed from the training tone by either one or two octaves. The same frequencies and intensities were used for this group as for the same tone group. Half of the *Ss* in the stimulus generalization group were trained to a tone lower than the test tone and half were trained to a tone higher than the test tone.

After the *Ss* came into the experimental room, *E* explained to them that they were to serve in a conditioning experiment. They were then told that conditioning would be accomplished by pairing a shock with a tone, and that, after a while, they would develop a tendency to respond to the tone alone. They were asked neither actively to try to respond to the tone, nor to inhibit the response if they felt like making it. They were told that the experiment would begin with a test of threshold to two tones, followed by a test to determine what level of shock was to be used.

Rough threshold measurements for two tones were made by the simple clinical method of lowering the intensity of the tone until *S* reported that he could no longer hear it, and then increasing the intensity until he reported that it could be heard. Since intensity was controlled in steps of five decibels, there was usually no difference in the up and down threshold. The intensity level of the tone that was used in the experiment was 50 decibels above the threshold.

Shock intensity varied from *S* to *S*. It was determined by stepping up the level of the shock until *S* reported that it was as strong as he desired to take. Occasionally, if the latent time of the shock reaction was long, *E* increased the shock during the experiment.

Following these various tests, the standard procedure was then begun. This procedure consisted of:

1. Three presentations of the tone alone. The particular tone used was of the frequency that would be employed in the final test. If any *Ss* responded to the tone alone at this time, they were discarded from the experimental group. Only one *S* had to be eliminated from the group for this reason.

2. Three shocks alone.

3. One hundred thirty paired stimulations. The time intervals between these stimulations varied from five to 35 sec. Four *Ss* out of 63 did not show any conditioned responses as measured by anticipatory reactions during this training, and these *Ss* were not continued on the experiment.

4. *E* entered the *S*'s room and told him: "Now, I want to turn your hand over so that the same finger is on the electrode, but your palm is up." The hand was then turned over, and the recording contacts adjusted to the new position. If *S* asked any questions at this time, *E* replied that he would answer them later.

5. Thirty sec. after *E* returned to the experimental room, he began to present the tone alone. This was presented for 10 trials. Time intervals between presentation varied, and were the same for all *Ss*.

The transfer test procedure consisted of 10 stimulations of the tone *alone*. For 27 *Ss* a tone the same as that of the training situation was used. For 16 *Ss* a tone differing by one octave from the training tone was used. For 16 *Ss* a tone differing by two octaves was used. The time intervals between stimulations varied in prearranged random order from five to 35 sec.

Results

1. *General results.*—The results of the study are presented in Table I. Column 1 identifies the groups and column 2 gives the number of *Ss* in each group. The next three columns indicate how the *Ss*

TABLE I

Showing the Number of *Ss* Who Failed to Respond, or Gave Specific or Generalized Responses When the Hand Was Turned Over

Group	Number	No Response	Generalized Response	Specific Response
Generalized tone	32	10	22	0
Same tone	27	2	21	4

reacted when the hand was turned over. Column 3 shows the number of *S*s who failed to respond at all; column 4 shows the number who gave flexor generalized responses; and column 5 lists the number who continued to make extensor or specific responses.

2. *Comparison of the frequency of response in the 'new' situation.*—If for the moment we disregard the nature of the response, that is whether it is generalized or specific, some tendency is shown for the group which was tested to the same tone to show a greater strength of conditioning than did the group tested to different tones. Twenty-five out of 27 in the same group responded, and 22 out of 32 in the stimulus generalization group responded. This produces a chi-squared value of 5.1670, which is significant between the two and one percent levels.

3. *Comparison of the frequency of generalized and specific responses in the two groups.*—The data of greatest interest in the present study are those which present a comparison of the tendency to make specific or generalized responses in the two groups. These comparisons are available from the material presented in Table I. Twenty-five of the *S*s who were tested to the same tone responded. Of this group, four gave specific responses—that is, made a response that in muscular terms was the same that they had been making during the original training—and 21 *S*s reacted with a generalized response—that is, changed the form of the response from an extensor movement to a flexor movement. In the group which was tested to a different tone, 22 *S*s responded in the transfer situation. Everyone of these *S*s made the generalized response.

A comparison of the two groups produces a chi-squared value of 3.8477. Such a value as this would occur by chance just a little less than five times in 100. One may state with a moderate degree of confidence that the difference in behavior of the two groups was not due to chance variation.

Discussion

The statistical evaluation of the experiment indicates some trend for the group tested to the same tone to be more likely to give a specific response to the conditioned stimulus than do the groups who are tested to a different tone. Assuming this is not due to sampling error, it would seem to represent some real functional relationship between response generalization and stimulus generalization. It should be remarked that the level of probability is modified somewhat when Yates' correction is employed. When this is done a *P*-value of about eight percent is obtained. Since this correction is arbitrary and of somewhat dubious value, the slight shift in *P*-value does not warrant a disregarding of the results of the experiment.

The results of the experiment seem to be related to the type of stereotyped behavior of which Guthrie's puzzle box cats are example (**2**). He reports that after some training in opening the box by contacting a pole, a marked disturbance in the behavior of the animal may be found when the pole is moved to a new location. This may be considered as a failure of the occurrence of response generalization. From a more positive point of view it is described by Guthrie as arising from the fact that the animals have learned a series of particular movements rather than acts. The present experiment would seem to indicate that whether an act or a movement is given to a stimulus situation is dependent upon the degree of similarity of the stimulus situation to the original training situation.

A crude analogy with the present experiment is found in the transposition experiment of Jackson, Stonex, Lane and Dominguez (**5**). These authors report that absolute responses are more likely to occur if considerable training is given before the transposition test, while a greater predominence of relative responses occur with less training on the original pair of stimuli. To support the analogy it would be necessary to assume that the degree of conditioning, the strength of the S-R connection, or magnitude of the ${}_{s}H_{R}$ was greater when the same tone was employed than when other tones were used. This assumption could be justified not only on the basis of internal evidence from the present experiment (the tendency to respond at all in the new situation was greater for the same tone group), but also from experiments on stimulus generalization (**1**, **3**, **4**). The analogy between the transposition experiment and the present one, however, is not close; for the *form of the response remains the same* in the transposition experiments, but the stimulus to which it is made either varies (as in the relational response) or remains identical (as in the absolute response). In the present experiment the significant variable is the modification or lack of modification of the form of the response itself.

In summary it would appear that response generalization may be a function of stimulus generalization, and that the introduction of new perceptual factors in the situation is not as likely to result in a modification of the response if the stimulus remains the same as it is if the stimulus is somewhat modified. It suggests that the organism learns to make a certain response to a certain stimulus and that the response persists in the presence of this stimulus, even though it is apparently nonadaptive, and even though another kind of response would seem to be more appropriate. The present experiment cannot indicate whether this result arises from the fact of difference in the nature of the stimulus *per se* or because the change in stimuli results in shifting to an S-R connection which is weaker in strength than the original S-R connection.

Summary

Two groups of *Ss* were conditioned to respond to a tone by making extensor movements of the finger. The hand was then turned over so that a flexor movement would be required to remove the finger from the electrode. Both groups were then tested to the conditioned stimulus alone. For the control group of 27 *Ss* the same tone was used, and for the experimental group of 32 *Ss* a different tone was used. Twenty-five *Ss* in the control, and 22 *Ss* in the experimental group gave conditioned responses in the new situation. All *Ss* of the experimental group gave generalized responses (made flexor movements). In the control group, four of the 25 responding *Ss* gave *specific* (extensor) responses in this situation. Thus, the results suggest some trend for response generalization, or the tendency for an act rather than a movement to become attached to a stimulus, to be a function of stimulus generalization.

(Manuscript received July 14, 1947)

References

1. Bass, M. J., & Hull, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17**, 47–65.
2. Guthrie, E. R., & Horton, G. P. *Cats in a puzzle box.* N. Y.: Rinehart, 1946.
3. Hovland, C. I. The generalization of conditioned responses. I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
4. Hull, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.
5. Jackson, T. A., Stonex, E., Lane, E., & Dominguez, K. Studies in the transposition of learning by children. I. Relative vs. absolute response as a function of amount of training. *J. exp. Psychol.*, 1938, **23**, 578–600.
6. Wickens, D. D. The transference of conditioned excitation and conditioned inhibition from one muscle group to the antagonistic muscle group. *J. exp. Psychol.*, 1937, **22**, 101–123.
7. Wickens, D. D. The simultaneous transfer of conditioned excitation and conditioned inhibition. *J. exp. Psychol.*, 1939, **24**, 332–338.

MEASUREMENT OF ACCOMMODATION AND CONVERGENCE TIME AS PART OF A COMPLEX VISUAL ADJUSTMENT [1, 2]

BY ROLAND C. TRAVIS

Laboratory of Sensory Psychology and Physiology, Tufts College

A number of experimental studies on accommodative facility were made during and shortly after World War I with special reference to pilot fitness. This same type of impetus in military situations toward placing great demands upon speed and precision of perceptual-motor adjustments, especially in the handling of aircraft, arose again during World War II and is still a very pertinent military problem. Ocular functions play a dominant role in the control of aircraft especially in take-offs and in landing. Important among the various ocular components essential in flying an airplane are the accommodative and convergent functions in successively focusing from the instrument panel to the take-off strip and from the take-off strip back to the instrument panel. The duration of this over-all performance, actual time for accommodation and convergence, eye-movement time, perceptual speed, and reaction time in making appropriate response, are crucial in successfully flying fast planes.

The author is in full agreement with Boring (**2**) who suggested that visual perception should be measured in large units rather than to attempt to isolate and to measure specific visual components. Thus, a technique to measure speed of accommodation and convergence ought to include measures of visual acuity, time for short eye-movements in refixation, time for recognition of the stimulus, and time to make an appropriate response, as well as a means of isolating the time for accommodation and convergence. The measurement of speed in several visual components combined would have a more realistic bearing on occupations demanding good over-all visual performance than would tests of specific components.

Historical Orientation

The importance of speed in making correct visual perceptions in relation to distance was emphasized by Ferree and Rand in 1918 (**3**). More recently

[1] Part of the experimental data given in this report were obtained in the Psychological Laboratory of Stanford University in conjunction with the Army Air Base at Fresno and the Naval Air Station at Moffett Field, California.

[2] This work is supported by the Psychology Branch, Aero Medical Laboratory, Engineering Division, Wright Field, Dayton, Ohio, under Contract No. W33-038 ac-14559.

these authors (4, 5) developed a technique that presented three test-objects at three different distances from the observer. The variables purported to be tested independently were relative distance of objects, their lateral separations, their illuminations, times of exposure, times between successive exposures, and rotary position of the test letters (E's). The *S*'s task was to report the angular position of the E's and their relative distances. The authors recommended the device as applicable to testing prospective pilots, to measuring effects of fatigue and for training in oculomotor and accommodative facility. They also suggested that this technique would be useful as a test for night flying by controlling the illumination on the test letters. Apparently no attempt was made to study the response-time of the subject in reporting the stimulus, nor apparently did they consider that appropriate response was an important part of the total performance.

Tefft and Stark (16) pointed out that accommodation time as such was not too important in pilot fitness as a separate measure.

Robertson (12) found a definite correlation between accommodation facility and age after thirty years.

Using two similar fall-tachistoscopes, Banister and Pollock (1) found individual differences in the time for recognizing letters while focusing from a distance of 41.5 cm. to a distance of 610 cm. Diurnal variations and age differences were found. Effects due to practice were not mentioned.

By using an instrument which controlled size and distance, Grant (7) found among other results that accommodation and convergence are practically inseparable functions. That is, when accommodation was demanded by an appropriate visual response to a stimulus, binocularly or monocularly convergence also occurred; contrariwise, when convergence was called into play in a special experimental set-up without a need for accommodation, the accommodative adjustment was nevertheless made. Furthermore, increasing accommodation and convergence not only constituted an adjustment to decreasing distance, but also reduced the apparent size of the visual image. Grant's results contradicted the earlier conclusion of Swenson (15) that accommodation and convergence are quite separable processes under appropriate stimulation.

In a study of speed of monocular accommodation, Leukart (8) found average times of 0.33 sec. and 0.37 sec. for near to far (1 m. and 4 m.) and far to near respectively for 14 *S*s. No significant differences between the right eye and left eye were found. For the four different situations, accommodation for right eye alone, accommodation for left eye alone, both eyes near to far, and both eyes far to near, individual scores ranged from .07 sec. to .70 sec. The means for the total scores for the high seven cases and the low seven cases were .45 sec. and .22 sec., respectively. Leukart avoided the time required in either a manual or verbal response to the near and far stimuli. A fixation stimulus and either the near or the far stimulus were presented in mirrors arranged on a rotation wheel in a manner which permitted constant fixation in one direction. The constant stimulus method was used, and the time between the fixation stimulus and the second stimulus demanding accommodative adjustment was systematically varied in preliminary trials to secure a rough measure of the range for each subject

before the main series. Each measurement was taken separately and the response-time did not enter into the accommodative adjustment time. This is one of the more important experimental studies in the measurement of accommodative speed as an isolated function.

Apparatus and Procedure

In the present experimental study, a specially designed drop-frame tachistoscope which could be operated by either *E* or *S* was used to present the near and far stimuli. Broken circles with the open segment near one of the numbers 1, 2, 3 or 4 served as discriminating stimuli for both the near and far distances. Fig. 1 shows the arrangement of the numbers around the

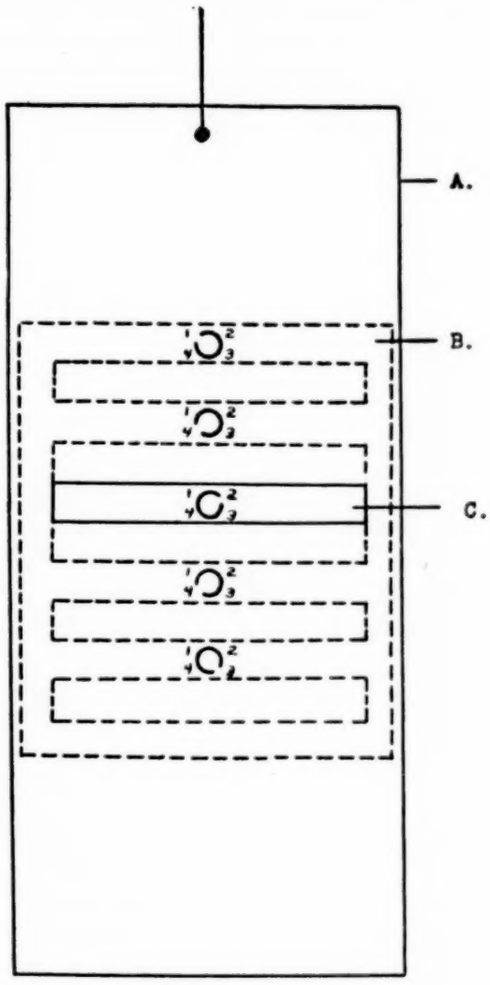

Fig. 1. Sketch showing relation of shutter screen to near-stimuli card

broken circle. The diameter of the broken circles subtended 0.6 degree and the breaks subtended 0.066 degree of visual angle. The diameter of the circles approximates the visual angle of 12-point typewriter print at 11 in. The numbers to indicate position of the break subtended a visual angle of 0.33 degree, and each number was always in the same position for all stimuli. The near-stimuli were 22 in. from the eyes and were placed on rectangular segments (½ by 4 in.) which alternated with rectangular apertures (½ by 4 in.) in a white cardboard frame (Fig. 1, B). A shutter-screen (Fig. 1, A) directly in front of this cardboard with a rectangular aperture (½ by 4 in.) (Fig. 1, C) slid up sharply whenever the lever was moved forward or backward. Each movement of the shutter-screen covered a distance of one-half in. (1.3 degrees) as the near and far stimuli were exposed.

The lever mechanism operated on the principle of the staggered catch drop-frame tachistoscope. Each movement of the lever permitted the weight to drop only one notch at a time. The weight was attached to the shutter-screen by a cable over two pulleys. By a simple movement of the lever forward or backward the *S* could expose near stimuli in succession, far stimuli in succession, alternately near and far, or in some other combination depending on the pattern of the near stimuli cards.

The far stimuli were mounted on cards 9 by 14 in. on a vertical white background 11 feet high and 43 feet from the *S*. This distance proved to be convenient in making the near and far stimuli comparable in size, in using the near distance of 22 in. and in using the exposure aperture of ½ in. in height. In successive exposures of the stimuli a short upward eye-movement of 1.3 degrees occurred for each exposure. The *S*'s head was supported on a head rest. The near stimuli were illuminated by an indirect lighting fixture, the far stimuli by a bank of direct lighting fixtures. Preliminary studies were made to obtain a phenomenological similarity in illumination of the near and far stimuli at approximately 10 foot-lamberts.

In a typical experimental session the lowest far stimulus was exposed and the *S* was instructed to focus his gaze on this stimulus. At the signal "go" the *S* gave a quick jerk backward on the lever and exposed the first near stimulus. As soon as he focused on the near stimulus and reported the number near the break in the circle he moved the lever quickly forward and exposed the second far stimulus. This continued until a series of 10 changes had been made. A short rest followed while stimulus cards were being changed.

Each series of 10 changes was timed in hundredths of a sec. by an electric stop clock which was automatically started and stopped at the beginning and ending of each series.

The positions of the break in the circles for both the near and far stimuli could be varied as desired.

Either the far or the near stimuli could be exposed successively in such a manner as to avoid accommodative adjustment. This provided a means of approximating more closely the actual accommodation and convergence time by subtracting the average times for the near stimuli responses and the far stimuli responses from the average times for the successive changes from far to near and from near to far.

Experimental Results

Many preliminary trials were necessary in calibrating the illumination, securing an optimal size of stimuli including the break in the circles and the numbers near the breaks, determining an optimal extent of eye movement between fixations, and securing a simple, positive and reliable self-pacing mechanism for presenting successive stimuli.

The near distance of 22 in. was a close approximation to the distance from the eyes of the pilot to the instrument panel in small aircraft. The far distance of 43 ft. was well beyond the commonly found accommodation limit of 20 ft. The size of the break in the circles, subtending 0.066 degree of visual angle, proved to be sufficiently small to demand accurate accommodation to recognize its rotary position, yet no *S* had difficulty in actually seeing the break.

Each *S* was given a warm-up trial to become acquainted with the task before the regular trials began. It was obvious that considerable learning would take place in the facility of performing this over-all task. Thus, the series of trials for the different experimental situations were staggered in each experimental session to control the influence of practice. In order to bring the experimental session within one hour duration two series of 10 successive fixations were given for each of the five situations. The order of the series was as follows: left eye refixating near and far stimuli—trial one; right eye refixating near and far stimuli—trial two; both eyes refixating near stimuli only—trial three; both eyes refixating far stimuli only—trial four; both eyes refixating near and far stimuli—trial five. These series were repeated for the second half of each experimental session.

Fifty college students ranging in age from 17 to 26 years took part in the first study. In view of the fact that there were no reliable sex differences, the results from both men and women will be presented together.

Table I gives the means, standard deviations, ranges and relevant reliability coefficients for the five experimental situations. Several important points are evident from a careful perusal of Table I. The mean times are significantly lower for the second trial for each situation than those for the first trial, which indicates the effect of practice. In general the range is more restricted on the second trials.

TABLE I

NUMBER OF CASES, MEANS, STANDARD DEVIATIONS, RANGES, RELIABILITY COEFFICIENTS FOR BOTH EYES, LEFT EYE, RIGHT EYE, NEAR STIMULI ONLY, AND FAR STIMULI ONLY

	N	Mean (Sec.)	S.D. (Sec.)	Range (Sec.)	Reliability Coefficient Pearson *r*	
B_1	50	1.13	.23	.80–1.89	.89	
B_2	50	1.00	.18	.72–1.52		
L_1	50	1.32	.30	.70–2.30	.81	.76
L_2	50	1.15	.30	.70–2.20		
R_1	50	1.23	.33	.70–2.30	.86	
R_2	50	1.16	.35	.77–2.90		
N_1	50	.96	.17	.59–1.45	.92	.84
N_2	50	.84	.14	.50–1.30		
F_1	50	.86	.22	.53–1.78	.91	
F_2	50	.82	.22	.50–1.50		
B–NF	50	.20	.11	.05– .58		

r for Trial 1 and Trial 2, entire series = .92 (N = 250).

The total average time for binocular refixation of near and far stimuli in succession was 1.06 sec. (This is the average of the means of B_1 and B_2 in Table I.) The difference between the total averages of binocular and monocular refixation of near and far stimuli was 0.15 sec. with a critical ratio of 2.1, which indicates only a slight superiority of both eyes over either one alone. A small insignificant difference between the mean times of the right and left eyes is evident.

The combined average times for refixating near stimuli only and far stimuli only were subtracted from the average times of refixating alternately near and far stimuli. This difference was 0.20 sec. and was statistically significant with a critical ratio of 10.5. This value of 0.20 sec. is a close approximation to the actual average time for this type of accommodative adjustment.

All these results show great individual differences, as will be noted in the standard deviations and ranges. The reliability coeffecients indicate a fairly high consistency on trials appearing close together.

To study the effect of practice, four of the *Ss* in the previous study were given intensive training using the same procedure over a

period of 10 days. Each S made 50 responses to near and far stimuli every other day. These results are shown in Fig. 2 and indicate a marked practice effect during these five trials. All four Ss showed a marked decrease in average reaction time and a marked decrease in variability as the trials progressed. When the actual time for accommodation and convergence was isolated by subtracting the average times for near and far stimuli only from the average times of successive refixations of near and far stimuli, very little change was found during the five trials. This would indicate that most of the practice effect oc-

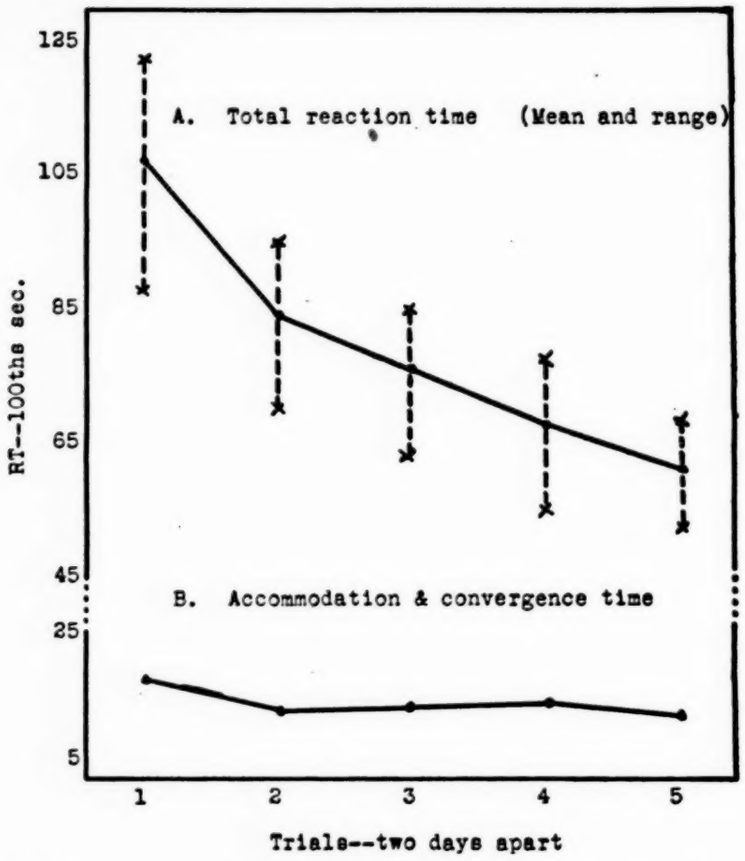

FIG. 2. Practice effect in response to refixating near and far stimuli. The mean for each trial represents 200 refixations for each of four Ss.

curred in components other than accommodation and convergence in this over-all performance. The course of learning during one experimental session is shown in Fig. 3 for 42 college students. Each trial represents 10 refixations of near and far stimuli with rest periods of two minutes between trials. As will be seen, considerable learning takes place from the first to the fifth trial, without much change in variability as indicated by the standard deviations. The procedure for this group of Ss was to fixate successively the near and far stimuli with both eyes; otherwise the procedure was the same as for the other two groups.

In order to analyze further visual acuity, perceptual speed, and motor speed in this complex task of responding to near and far stimuli, three additional tests along with the test to near and far stimuli were administered to 42 college students. One of these tests involved speed of coordinated movement of the arm and hand in a reach and turn performance in a certain serial order. The perceptual test was the Whipple Letter Cancellation Test. Visual acuity was measured by a standard Snellen Chart and the scores were converted into percentage scores. Intercorrelation coefficients for the four tests were calculated along with reliability coefficients for each test. Some idea of the contribution of visual acuity, perceptual speed and motor speed to the total performance should be obtained by the correlation method.

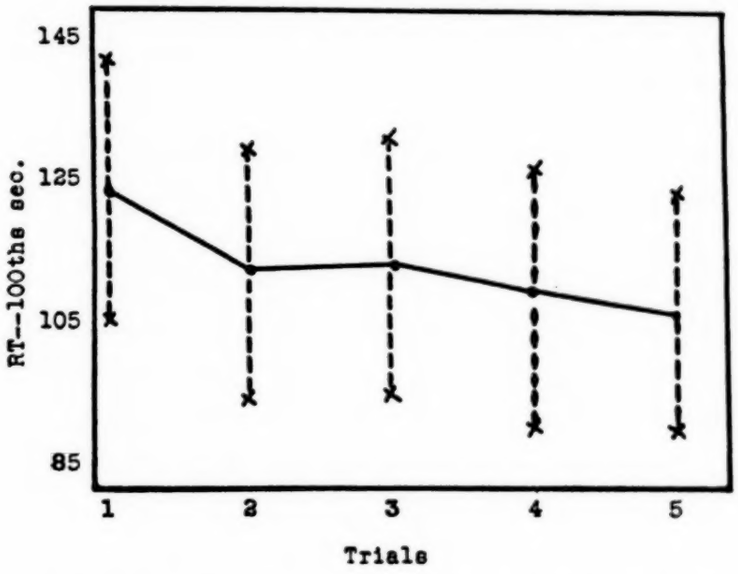

FIG. 3. Means and SD's for 42 college students for five trials with 10 fixations of near and far stimuli for each trial—two min. rest periods between trials

Table II gives the intercorrelation and reliability coefficients for the various tests. In view of the fact that speed in motor and perceptual tests increases with practice, the average of the scores for the first two trials was used in the correlations.

From the intercorrelations in Table II it is evident that visual acuity, perceptual speed as measured by a cancellation test, and motor speed as measured by a reach and turn performance, make a definite contribution to the total task of perceiving and responding to near and far stimuli. Visual acuity as measured by the Snellen Chart seems to have no relation to perceptual speed or motor speed. The negative correlation between visual acuity and accommodation follows from the fact that decreasing accommodation times were associated with increasing visual acuity scores.

TABLE II

Intercorrelation and reliability coefficients for four tests, (a) visual acuity (corrected vision), (b) motor speed, (c) perceptual speed, and (d) successively refixating and reporting on near and far stimuli (this includes visual acuity, recognition of position of break in circle, verbally reporting its position and manually moving the lever for the next exposure as well as accommodation and convergence).

Intercorrelations (N = 42)		Pearson r	PE_r
Accom. and Conver.	Vis. Acuity	−.56	.07
Accom. and Conver.	Per. Speed	.39	.09
Accom. and Conver.	Motor Speed	.48	.08
Per. Speed	Motor Speed	.44	.09
Vis. Acuity	Per. Speed	.19	.10
Vis. Acuity	Motor Speed	.08	.10
Accom. and Conver.	Age (17–26)	.00	

Reliability, Trial 1 vs. Trial 2 (N = 42)	Pearson r	PE_r
Accom. and Conver.	.90	.02
Vis. Acuity	.85	.03
Per. Speed	.86	.03
Motor Speed	.88	.02

Fisher's test of significance of r with 42 cases and two variables requires 0.30 for the five percent level and 0.39 for the one percent level.

The correlation between age for this group and the over-all reaction time was zero.

Summary and Conclusions

The speed of monocular and binocular accommodation and convergence was measured as part of a complex visual performance by use of a specially designed drop-frame tachistoscope.

The major conclusions may be summarized as follows:

1. The total average time for binocular vision in refixating near and far stimuli successively, a task which includes visual acuity, recognition of the stimulus, a verbal report, a manual movement of a lever, accommodation and convergence was 1.06 sec. for 50 Ss.

2. The total average time for accommodation and convergence alone in binocular refixation of near and far stimuli was 0.20 sec. for 50 Ss.

3. Binocular refixation was slightly superior to monocular refixation of near and far stimuli. The difference between the right and left eyes was not significant.

4. Considerable consistent practice effect occurred in successive trials and during successive days for the over-all task, but little or no practice effect for accommodation was found.

5. Large individual differences were found both in times for over-all performance and in actual times for accommodation and convergence.

6. Intercorrelations between the over-all performance of refixating near and far stimuli, visual acuity, perceptual speed and motor speed

indicated that the latter three, components contribute definitely to this over-all accommodation and convergence task.

(Manuscript received July 25, 1947)

References

1. Banister, H., & Pollock, K. G. The accommodation time of the eye. *Brit. J. Psych.*, 1928–1929, **19**, 394–396.
2. Boring, E. G. Psychology of perception: its importance in the war effort. *Amer. J. Psychol.*, 1942, **55**, 423–435.
3. Ferree, C. E., & Rand, G. The inertia of the eye for clear seeing at different distances. *Amer. J. Ophthal.*, 1918, **1**, 764–776.
4. Ferree, C. E., & Rand, G. A multiple exposure tachistoscope. *J. exp. Psychol.*, 1937, **21**, 240–259.
5. Ferree, C. E., & Rand, G. Pilot fitness: a safety factor in aviation. *Brit. J. Ophthal.*, 1940, **24**, 581–597.
6. Goodall, E. B. The speed of accommodation. *Air Serv. Inform. Circ.*, 1920, **1** (3), 63–69.
7. Grant, V. W. Accommodation and convergence in visual space perception. *J. exp. Psychol.*, 1942, **31**, 89–104.
8. Leukart, R. H. The speed of monocular accommodation. *J. exp. Psychol.*, 1939, **25**, 257–270.
9. Morgan, M. W. The clinical aspects of accommodation and convergence. *Amer. J. Optom.*, 1944, **21**, 301–313.
10. Robertson, C. J. Measurement of the speed of adjustment of the eye to near and far vision. *U. S. Nav. Med. Bull.*, 1934, **32**, 275–283.
11. Robertson, C. J. A comparative study of the measurement of the speed of adjustment of the eye for near and far vision. *U. S. Nav. Med. Bull.*, 1935, **33**, 187–205.
12. Robertson, C. J. Measurement of speed of adjustment of the eye to near and far vision. *Arch. Ophthal.*, 1935, **14**, 82–89.
13. Robertson, C. J. Measurement of speed of adjustment of eye to near and far vision: A further study. *Arch. Ophthal.*, 1936, **15**, 423–434.
14. Robertson, C. J. Effect of fatigue on the adjustment of the eye to near and far vision. *Arch. Ophthal.*, 1937, **17**, 859–876.
15. Swenson, H. The relative influence of accommodation and convergence in the judgment of distance. *J. gen. Psychol.*, 1932, **7**, 360–380.
16. Tefft, L. E., & Stark, E. K. The speed of accommodation as a practical test for fliers. *Amer. J. Ophthal.*, 1922, **5**, 339–342.

A BEHAVIORAL ANALYSIS OF DEGREE OF REINFORCEMENT AND EASE OF SHIFTING TO NEW RESPONSES IN A WEIGL-TYPE CARD-SORTING PROBLEM[1]

BY DAVID A. GRANT AND ESTA A. BERG

University of Wisconsin

Introduction

The object of this experiment was to discover the functional relationship between degree of reinforcement of individual correct responses and the ease with which the *S* can shift from these responses to others in a multiple solution card-sorting problem. The problem which was used was the University of Wisconsin Card-Sorting Test. An earlier form of this abstract reasoning test has been described (**1**). It combined many of the features of the Weigl (**14**), Goldstein-Scheerer (**4**), and the Vigotsky tests (**13, 6, 7**), but permitted objective, quantitative scoring.

In a task of this sort, reinforcement of correct responses might be expected to serve a dual function. On the one hand, as emphasized by Guthrie (**5**), increased reinforcement of a particular stimulus-response relationship has usually been found to facilitate learning of that particular relationship. Furthermore, increased reinforcement would be expected to make it more difficult to shift from the original reinforced response to others (**8**). On the other hand, it has been shown that reinforcement of single responses is sufficient to cause the *S* to respond selectively to overall sequential characteristics of the situation. Examples of response to overall characteristics of the experiment are found in the temporal maze (**12**), conditioning (**9, 10, 11**), and maze-running (**2**). The present investigation was designed to reveal how these two learning principles would operate in this particular type of abstract reasoning problem.

Procedure

The University of Wisconsin Card-Sorting Test differed slightly from the earlier form. The materials consisted of a pack of four *stimulus cards* and 64 *response cards* which were devised so that each card contained from one to four identical figures of a single color. Four kinds of figures were used: stars, crosses, triangles, and circles. Four different colors were used: red, yellow, green, and blue. A single card might have four red stars, or two green circles, or any of the 64 possible combinations of colors, numbers, and forms. Each card could then be sorted

[1] This research was supported in part by the Research Committee of the Graduate School of the University of Wisconsin from special funds voted by the State Legialature for 1946–47.

or categorized according to the color, the number, or the form of the figure. The four stimulus cards were: *first*, one red triangle; *second*, two green stars; *third*, three yellow crosses; and *fourth*, four blue circles.

The four stimulus cards were placed before the *S* from left to right in the order named. The *S* was given the pack of response cards and was instructed in the following manner: "I want you to put these cards into four groups, underneath the ones lying on the table. I will tell you whether you are "right" or "wrong."

The initial 'correct' sorting category was arbitrarily determined in advance to be *color*. As the *S* sorted the response cards he was informed whether he was 'right' or 'wrong.' As soon as the *S* made a certain number of consecutive correct responses (reinforcing or confirming trials), the *E shifted the problem with no explanation to S and began to call the number classifications 'right' and all others, including color, 'wrong.'* In this way, the 'correct' classification or category was later shifted from number to form, then back to number, then to color, and finally to form. The *S*'s only cue to the shifts was in the *E*'s 'rights' or 'wrongs.' If the *S* used the whole pack of cards before the six successive sorting categories were completed, the cards were reassembled and handed back to him. The cards were always presented in a standardized order so that no like form, color, or number followed each other. The experiment was considered completed after the *S* had finished the six categories or used more than 64 cards in a single category.

Seven groups of *S*s were run with 3, 4, 5, 6, 7, 8, and 10 reinforcing trials (the number of consecutive correct responses), respectively, before each shift. About 20 *S*s completed the test in each group.[2] All were men and women students who volunteered from the elementary psychology classes at the University of Wisconsin. Assignment to groups was determined on the basis of order of the *S*'s appearance for the experiment.

Results

Each response or classification of a single card was scored as 'correct' or 'incorrect' according to the arbitrary classifications set up by the *E* and shifted by him as described above. In addition, the *S*'s apparent criterion of classification was recorded when this was possible. For purposes of analysis, responses were classified as: (A) correct; (B) errors; (C) perseverative responses; and (D) non-perseverative errors. A perseverative response was defined as a response to a new or shifted category which would have been correct for the immediately preceding category.[3] Perseverative responses were usually errors. They represented failures to shift solution. They tended to prevent fresh modes of attack. Non-perseverative errors were computed by subtracting the number of perseverative responses from the total number of errors for each *S* for each category. Non-perseverative errors represented attempts to shift the mode of attack.

The means and variances for the total number of errors per *S*, total number of perseverative responses per *S*, and total number of non-perseverative errors are given for each group in Table I. These data are summarized graphically in Fig. 1, where mean total errors, mean total perseverative responses, and mean total non-perseverative

[2] See Table I for exact numbers.

[3] Whenever an *ambiguous correct response* was both preceded and followed by a perseverative error, it was counted as a perseverative response. E.g., suppose that after shifting from *color* (C) to *form* (F) the responses are: C, C, C, CF, C, CNF, C, F, F, F, F, F. These 12 responses yield five errors, seven perseverative responses, and seven correct responses.

TABLE I

MEANS AND VARIANCES OF TOTAL ERRORS, PERSEVERATIVE RESPONSES, AND NON-PERSEVERATIVE ERRORS AS FUNCTIONS OF THE NUMBER OF REINFORCING TRIALS

Response Type		Number of Reinforcing Trials						
		3	4	5	6	7	8	10
		$N = 19$	$N = 22$	$N = 21$	$N = 20$	$N = 21$	$N = 21$	$N = 20$
Total Errors	M	29.63	31.86	24.95	23.85	21.95	22.71	18.60
	σ^2	365.49	342.85	215.00	180.33	173.57	176.20	127.74
Perseverative Responses	M	18.32	21.59	17.19	12.50	16.33	15.86	12.60
	σ^2	165.16	166.79	149.96	38.79	120.98	166.31	77.64
Non-perseverative Errors	M	8.95	9.41	3.95	7.55	3.95	4.86	3.10
	σ^2	97.42	85.61	13.85	65.85	29.38	37.07	11.83
Perseverative / Total Error	%	62	69	69	52	74	70	68

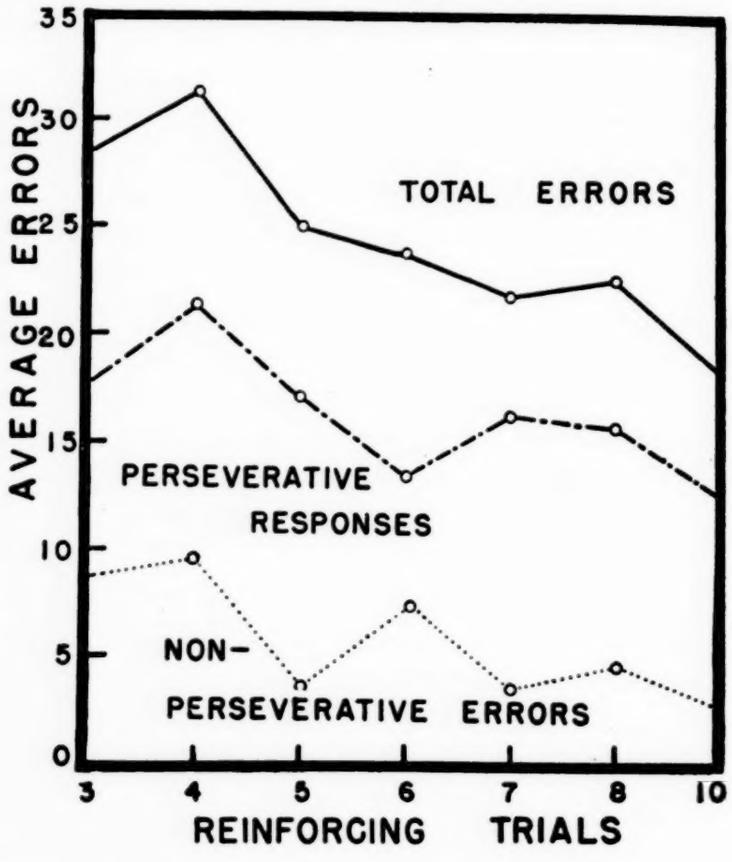

FIG. 1. Total errors, perseverative responses, and non-perseverative errors plotted against the number of reinforcing trials

errors are plotted against number of reinforcing trials. It will be observed that the dominant trend is downward; i.e., the average number of errors decreases with an increase in the number of reinforcing trials. Both the perseverative responses and the non-perseverative errors participate in the downward trend. Furthermore, the linear components of two of the three trends are statistically significant as shown by the Pearson product-moment correlation coefficients between the error scores and the number of reinforcing trials. The coefficients are: −0.25, −0.19, and −0.25 for total errors, perseverative responses, and non-perseverative errors, respectively. The two coefficients of −0.25, each based upon 142 degrees of freedom, are significant at the two percent confidence level (**3**, p. 42).

In the bottom line of Table I the perseverative responses are

TABLE II

MEANS, VARIANCES, AND SIGNIFICANCE OF DIFFERENCES IN RESPONSE SCORES BETWEEN LOW-REINFORCED (A) AND HIGH-REINFORCED (B) GROUPS AS A FUNCTION OF SUCCESSIVE SORTING CATEGORIES

Group			Categories					
			1	2	3	4	5	6
A 3, 4, and 5 Reinforcing Trials $N = 62$	Total Errors	M	2.86	5.32	6.32	2.73	6.90	4.71
		σ^2	14.12	31.19	32.61	9.65	81.73	34.43
	Perseverative Responses	M		3.39	5.84	2.42	5.10	2.35
		σ^2		15.66	34.91	9.40	49.15	6.42
	Non-perseverative Errors	M		2.03	0.81	0.31	1.97	2.37
		σ^2		8.71	7.61	1.63	11.42	23.75
B 6, 7, 8, and 10 Reinforcing Trials $N = 82$	Total Errors	M	2.50	5.21	5.23	2.60	3.29	2.96
		σ^2	12.88	26.73	39.05	6.22	15.21	10.21
	Perseverative Responses	M		3.59	4.56	2.10	2.63	1.61
		σ^2		11.68	27.72	5.33	13.84	2.46
	Non-perseverative Errors	M		1.54	0.78	0.49	0.70	1.35
		σ^2		15.37	7.37	1.71	3.02	5.42
t's for Differences Between Groups A and B	Total Errors		0.58	0.12	1.11	0.28	3.22**	2.28*
	Perseverative Responses			0.32	1.36	0.71	2.71**	2.14*
	Non-perseverative Errors			0.82	0.07	0.89	2.91**	1.65

* Significant at 5 percent confidence level
** Significant at 1 percent confidence level

expressed as percentages of the total number of errors. There is no pronounced upward trend in the percentages, which is contrary to the prediction that increased reinforcement would make shifting more difficult. Such trend as does exist is not statistically significant. Furthermore, the *median* percentages reverse the slight upward trend of the mean percentages in Table I. There is thus no evidence for increased perseveration, absolute or relative, following a greater number of reinforcement trials.

The chief effect of increased reinforcement of old solutions is to facilitate *new* solutions. How increased reinforcement facilitates new solutions is revealed by a more detailed analysis in terms of performance on successive categories. In Table II, the means and variances for total errors per *S*, perseverative responses per *S*, and

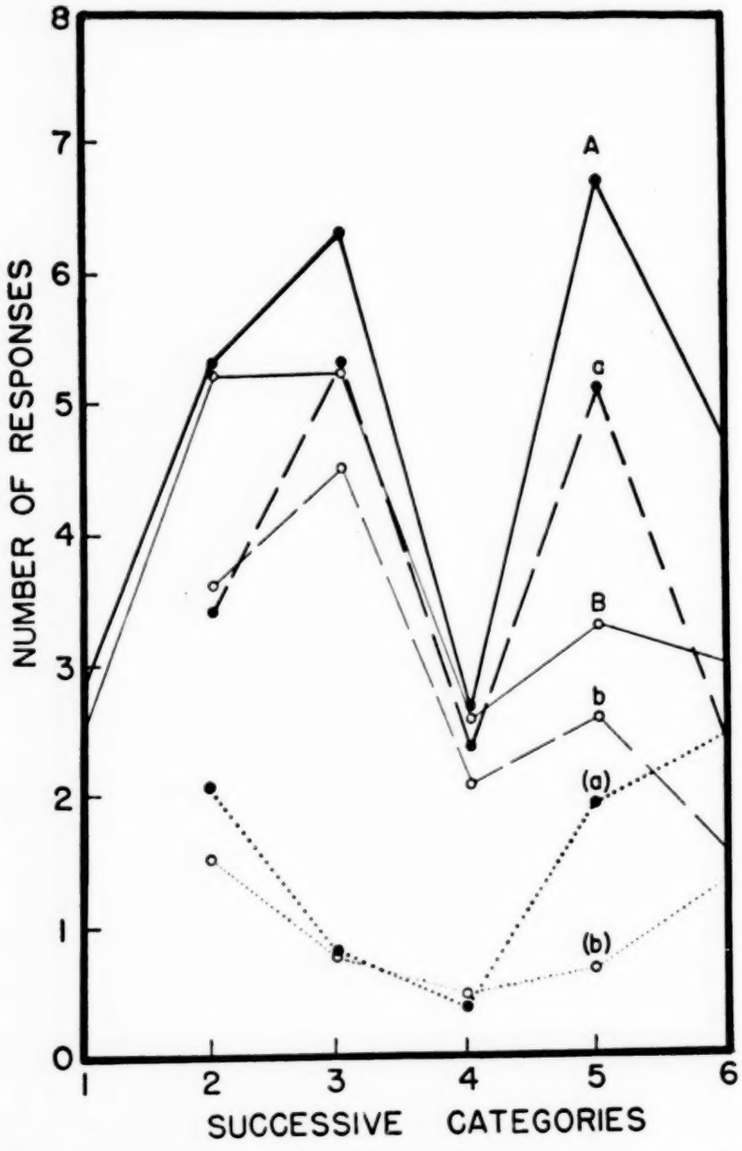

FIG. 2. Total errors, perseverative responses, and non-perseverative errors plotted for successive sorting categories with degree of reinforcement as a parameter

non-perseverative errors per *S* are given for each successive sorting category. The data of the three-, four-, and five-reinforcing-trial-groups are combined as Group A, and the data of the six-, seven-, eight-, and ten-reinforcing-trial groups are combined as Group B. The *t*'s for the differences between Group A and Group B in Total Errors, Perseverative Responses, and Non-perseverative Errors are given in the bottom rows of Table II. The data of Table II are plotted in Fig. 2, where the bold lines follow the means of Group A and the fine lines follow the means of Group B. Total errors are shown by solid lines, A and B. Perseverative responses are shown by the corresponding broken lines, a and b, and Non-perseverative errors are shown by the dotted lines (a) and (b).

From Table II and Fig. 2 it is apparent that the total error scores for Groups A and B remain nearly identical until the fifth category where the trends diverge sharply. This is borne out by the *t*-tests which are not significant, ranging from 0.12 to 1.11, on categories one to four,[4] but are statistically significant on categories five ($t = 3.22$) and six ($t = 2.28$)

The significance of the inter-group differences in perseverative responses and non-perseverative errors on category five show that both types of response are involved in the reduction of total errors. In other words, an increase in number of reinforcing trials reduces perseveration (i.e, makes it easier to shift) and reduces non-perseverative errors (i.e., reveals more clearly the possible alternative solutions.) Of the two types of errors the difference in perseverative responses remains significant on the sixth category, but the difference in non-perseverative errors is no longer significant because of increased variation between *S*s on this category.

Discussion and Summary

This experiment attempted a behavioral analysis of certain aspects of the reinforcement process in a complex problem-solving situation which involved some of the basic principles of the Weigl, the Vigotsky, and the Goldstein-Scheerer type problems. With this complex task, increasing the amount of reinforcement of original modes of response reduced the amount of perseveration of these responses when they suddenly became incorrect. At the same time, *S* was able to reach the 'new' correct solution more readily after he deserted the 'old' correct solution. The efficacy of increased reinforcement in bringing

[4] The sharp drop in all means on category four is characteristic of the 'key' that was used for the card-sorting test in this experiment: 1. color; 2. number; 3. form; 4. number; 5. color; and 6. form, were successively correct. On the basis of unpublished data, perseveration from category *two* (number) to category *four* (also number) seems to be the factor responsible for the drop.

about these changes was not immediately apparent but appeared clearly on the fifth category or *fourth shift of the correct mode of response.*

Interpretation of the role of increased reinforcement may be made in holistic terms. The abstract attitude is either fostered or made more effective by added reinforcement of the single correct responses. This is clearly indicated by the greater ease of shifting the mode of response from one aspect of the stimulus to another, the result found in the groups receiving more reinforcing trials. The fact that greater flexibility rather than greater rigidity of response occurs may be a function of the superior young adults who served as *S*s. In holistic terms, however, the greater flexibility may be interpreted in terms of the 'confirming' function of the reinforcing trials. Confirmation reduces the ambiguity of the situation, the complexity of which is so great that it requires several trials to produce closure.

The complex learning or problem-solution which takes place in this task may also be described in stimulus-response terms. Although the stimulus variables are novel, and have important temporal as well as spatial characteristics, the sort of process which takes place is by no means new to S-R formulation. Briefly, individual responses are reinforced or rewarded. But reinforcing a single response, *per se*, will not lead to solution. The *S* must learn to respond to *sequences* of stimuli; the discriminations necessary to solution extend over a series of trials. It is quite clear, as pointed out in the Introduction, that a *series* of reinforcements of isolated responses is adequate to establish the overall pattern of response in this and in many other experimental situations which have been brought into the psychological laboratory.

The quantitative nature of the various error scores in the Wisconsin Card-sorting Test permits a fairly detailed analysis of the role of reinforcement. The clear-cut reduction of perseveration following increased reinforcement shows that *S* learns to *shift* his mode of response, at least *he learns to reject quickly the formerly correct mode of response.* The fact that greater reinforcement reduces Non-perseverative errors shows that the *S* learns to *limit* his shifts to the predetermined 'correct' modes of response; *he becomes more selective in the way he shifts to other modes of response.* There is undoubtedly more to the learning process than has been described above, and it is not claimed that the process of solution of Weigl-, Vigotsky-, or Gelb-Goldstein-, Scheerer-type tests has been specified definitively. The present study, however, shows how analysis of the operation of important variables, such as single-trial-reinforcement, may be attempted in these complex tasks.

(Manuscript received July 11, 1947)

References

1. Berg, Esta A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948. (In press)
2. Brunswick, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
3. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver and Boyd, 1943.
4. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, 1–151.
5. Guthrie, E. R. *The psychology of learning.* New York: Harpers, 1935.
6. Hanfmann, Eugenia, & Kasanin, J. A method for the study of concept formation. *J. Psychol.*, 1936, **3**, 521–540.
7. Hanfmann, Eugenia, & Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, **27**, 3–113.
8. Hull, C. L. A functional interpretation of the conditioned reflex. *Psychol. Rev.*, 1929, **36**, 498–511.
9. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
10. Humphreys, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
11. Humphreys, L. G. Extinction of conditioned galvanic skin responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–75.
12. Hunter, W. S. The temporal maze and kinesthetic sensory processes in the white rat. *Psychobiol.*, 1920, **2**, 1–17.
13. Vigotsky, L. Thought in schizophrenia. *Arch. Neur. Psychiat.*, 1934, **31**, 1063–1077.
14. Weigl, E. On the psychology of the so-called processes of abstraction. *J. abnorm. soc. Psychol.*, 1941, **36**, 3–33.

SPATIAL ORIENTATION IN THE WHITE RAT*

BY H. C. WILCOXON

Yale University

AND

R. H. WATERS

College of Wooster

In a recent series of articles (**3, 4, 2, 5**) Tolman and his students have presented evidence in support of the theory of expectancy. According to this theory, when an animal learns a maze he learns a spatial orientation towards the goal and builds up an 'expectancy' of reward at that goal *place*. This theory is in opposition to the stimulus-response doctrine, which denies that any such mentalistic concept as expectancy need be introduced to explain behavior.

It is often difficult to distinguish between this theory of expectancy, or place learning, and the alternative stimulus-response theory, and it is even more difficult to design an experiment which will lend support to one of the theories and not to the other as well. There seems to be little doubt that an animal, after becoming familiar with a certain goal location as defined by certain differential cues at or near the goal location, can later exhibit a tendency to return to that location when started from a new position in the same environment. The recent series of articles on spatial learning from the California laboratory has served to illustrate this point clearly. These observations flatly contradict the older Watsonian chain-reflex theory of learning but by no means silence the newer and more flexible stimulus-response theories. As a consequence, it is believed that the legitimacy of the use of such expressions as 'expectancy' and 'spatial orientation' as terms descriptive of the factors underlying such behavior is still open to doubt.

No precise definition of spatial orientation is to be found in the studies reporting 'expectancy' behavior. Nevertheless it is clear that in order for the term to have legitimate descriptive status it must be defined in somewhat the following manner:

> Spatial orientation refers to that condition within an organism which enables the organism to make a response to a *spatial relationship* of objects in the environment, which *relationship* is determined solely by the *relative positions* of the objects in the environment and the organism itself.[1]

* This study was carried out at the University of Arkansas.

[1] Thus spatial orientation as used here differs from directional orientation as used by Dashiell (1) in that the former depends upon a perception of a spatial relationship derived at least in part

It is to be noted that the 'spatial learning' in rats recently demonstrated by Tolman *et al.* has not been shown to be dependent upon the rat's perception of a spatial relationship of objects per se for the reason that in each of the four studies published there have been present in the environment certain differential cues to which the rats conceivably could have become conditioned. That is, the rats may have made their responses towards discriminable objects rather than as a result of any perceived spatial relationship of objects. Consequently, 'spatial learning,' as used by Tolman, is not necessarily dependent upon spatial orientation as defined above. It is believed, however, that the interpretation of this type of learning in terms of expectancy assumes the rat's ability to perceive such a relationship and thereby gain a spatial orientation in the sense we have defined it. Hence, for additional support of Tolman's theory it would be desirable to show:

1. That the rat is capable of building up a spatial orientation which is clearly dependent upon a *spatial relationship.*

2. That the rat will develop an expectancy of reward at the place towards which he becomes oriented.

3. That the rat will demonstrate this spatial orientation and expectancy by taking a short-cut route to the place. With these three things in mind the following experiments were performed.

EXPERIMENT I

A. Animals

The animals used in both of the present experiments were albino rats from a local stock and were between 75 and 90 days old at the outset of training. In Experiment I 32 animals, 16 male and 16 female, with no previous maze experience were employed. The food used was Purina Dog Chow checkers and the animals were run and fed between 7:00 and 10:00 P.M. in both experiments. Water was in the home cages at all times.

B. Apparatus

The apparatus used in Experiment I was Maze I, the plan of which is shown in Fig. 1. This maze consisted of pathways elevated 36 in. from the floor with all paths made of two-in.-wide pine painted flat black. The starting pathways were curved to the left in an arc of about 300 degrees and were each eight feet long. The straight center path running from the food platform at F-1 to the food platform at F-2 was also eight feet long. On this path, four in. from each end, were wire frames from which were suspended black curtains eight in. high and 10 in. wide. Short-cut pathways, each four feet long, could be attached to either side of either curved starting pathway, extending to the food platforms.

One checker of dog chow was on each food platform at all times with a wooden block placed behind the curtain at the food platform which was incorrect for a given run.

This apparatus was surrounded by a curtain of unbleached muslin hanging from a circular frame of heavy gauge wire suspended six feet from the floor. The top of the circular curtain

from exteroceptive stimulation, whereas the latter was believed to be dependent upon intra-organic cues which served to keep the organism oriented once it started moving in a given direction.

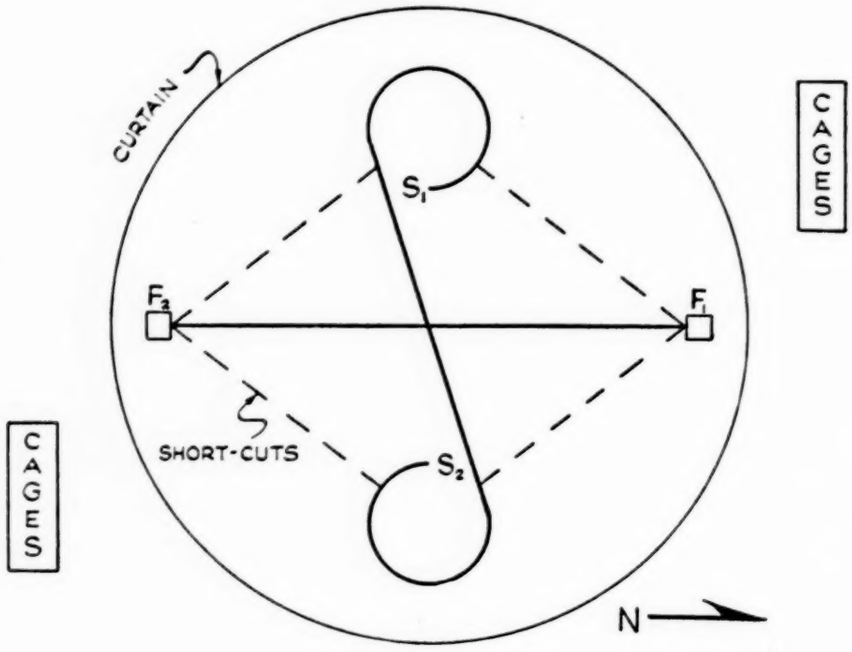

FIG. 1. This figure shows the floor plan of Maze I used in Experiment I. The dotted lines indicate the four possible short-cuts, one of which was added on the critical trial.

was covered with the same material and the floor of the room was painted flat black. The only illumination in the laboratory during the runs consisted of two shaded, 15-watt bulbs, one of which was placed directly over the center of the circle described by each starting pathway. This lighting arrangement made the curtain essentially a one-way vision screen. The entire apparatus was placed in a room 15 feet wide and 30 feet long, with the center pathway, F-1 F-2, running north and south.

The home cages were placed in positions symmetrical to the maze, as is shown in Fig. 1, since a previous investigation (unpublished) had indicated that auditory cues from the home cages would have to be controlled.[2] There were no other animals in the laboratory.

C. PROCEDURE

All animals were given identical training. Preliminary training was given on pathway F-1 F-2 on the first three days of the experiment. In this training the animals were fed on each of the two food platforms and given practice in running under the small black curtains to food.

On the fourth day the experimental training was begun with the curved starting pathways in position. Each animal was given four trials a day alternated in an ABBA order. Whenever an animal was started at S-1, the platform at F-2 was blocked. And conversely, when started at S-2, the platform at F-1 was blocked. Thus all of the animals were required to learn always to go left to obtain food. An error was recorded whenever a rat made a body length entry into either the pathway leading to the incorrect food platform or the pathway leading to the opposite starting position.

The training was continued until all of the animals either learned the problem to a criterion of 10 successive errorless runs, or developed stereotyped responses which seemed to indicate that they would never learn.

As the animals met the criterion of learning they were each given one critical trial on the maze with one of the short-cut pathways added. This critical trial was given on the same day the animal met the criterion. Half of the animals were tested to determine whether they would take a short-cut to the correct place, i.e., the platform at which they had been trained to 'expect'

[2] This confirms Ritchie's (2) finding that auditory cues from home cages are enough to support place learning.

food. These animals were designated the Experimental Group. The other half were tested to see whether they would take an identical short-cut pathway to the 'incorrect platform,' i.e., the platform at which they had never received food when started from the given position. These animals were designated the Control Group. Whether an animal was placed in the Control Group or in the Experimental Group depended upon when it met the learning criterion. The first animal to learn was placed in the Experimental Group, the second in the Control Group, the third in the Experimental, the fourth in the Control, etc. The starting position for the critical trial of any given animal was the same as it would have been on the next regular trial. When an animal in the Experimental Group was given its one critical trial, the only short-cut on the apparatus was the one leading from the given starting position to the correct place. Similarly, when an animal in the Control Group was given its one critical trial, the only short-cut on the apparatus was the one leading from the given starting position to the incorrect place.

Note that in each of the four possible short-cut setups an animal had to make a *right* turn off the starting pathway in order to get on the short-cut pathway, a response never rewarded in the training.

D. Results

Six of the animals failed to meet the criterion. All of these animals developed right-turning habits at the choice point, so that their performance consisted of at least one error on each trial after the habit became established.

The remaining 26 animals learned the maze to the criterion in an average of 7.9 trials, and half were placed in each group. Of the 13 animals placed in the Experimental Group, 12 took the short-cut presented to them, i.e., the short-cut to the correct food platform. Of the 13 animals placed in the Control Group, 10 took the short-cut presented to them, i.e., the one to the incorrect food platform. Of the four animals which learned the maze and were tested but did not take a short-cut, three proceeded without error over the old behavior route, and the fourth made a right-turning error at the old choice point.

E. Discussion

If significantly more animals had taken the short-cut to the correct place than to the incorrect place an interpretation in terms of spatial orientation would clearly have been indicated. It could have been argued that the animals had built up a spatial orientation which was dependent upon the spatial relationship of the starting position and the correct food platform, that they had learned to expect food at a place defined solely by this relationship, and that they had demonstrated these two facts by taking an entirely new pathway to the correct place and ignoring an identical one to the incorrect place. But almost equal numbers of animals took the two short-cuts, which renders the results equivocal for this fact is not evidence *against* the Tolman theory.

Consider the 26 animals which learned and were tested on the short-cuts. Twenty-two of them, or 85 percent, did take one of the short-cuts. In all cases, taking the short-cut involved the animal in

a right turn toward one of the food platforms, a response which had never been rewarded in the training. Three possibilities are offered in explanation of this result.

1. The animals might have taken the short-cut because it was a new stimulus which inhibited the old response and called forth some sort of exploratory behavior.

2. The animals might have taken the short-cut because they expected food behind the black curtain at the end of *either* short-cut pathway. However, if this were true it could be argued that the animals were responding to an object, i.e., a black curtain, and not to a spatial *relationship* of objects.

3. The animals might actually have built up a spatial orientation and expectancy of food at the correct place but did not demonstrate it clearly because of certain conflicting aspects of the experimental situation. The fact that on the critical trial each animal was faced with a new stimulus, viz., the short-cut, may have been important, as suggested above. And if a new stimulus could serve so effectively to inhibit the old response, it might also have inhibited any demonstration of spatial orientation and expectancy which, according to the Tolman theory, was an important factor in the old response. A better experiment would be one in which the short-cut, as a new stimulus, could not wield such a decisive influence. An apparatus which would allow the simultaneous presentation of two short-cuts, one to the correct place and the other to the incorrect place, would overcome this difficulty.

Each of the above possibilities appeared equally likely and suggested further research.

EXPERIMENT II

A. Introduction

In the preceding experiment the evidence indicated that the short-cut pathways, acting as new stimuli, might have elicited exploratory, rather than expectancy behavior. The present experiment was designed to overcome this difficulty. An apparatus was designed which allowed the simultaneous presentation of two short-cuts, so that the animal was required to make a choice between two new pathways, rather than between an old one and a new one as in the previous experiment. Thus any tendency toward exploration was automatically equalized as regards the pathway the animal might take on the critical trial.

B. Animals

Ten male rats with no previous maze experience were used.

C. Apparatus

The apparatus was the same elevated maze described in Experiment I with the exception that straight starting pathways 30 in. long were substituted for the curved ones used in the previous experiment (see Fig. 2).[3]

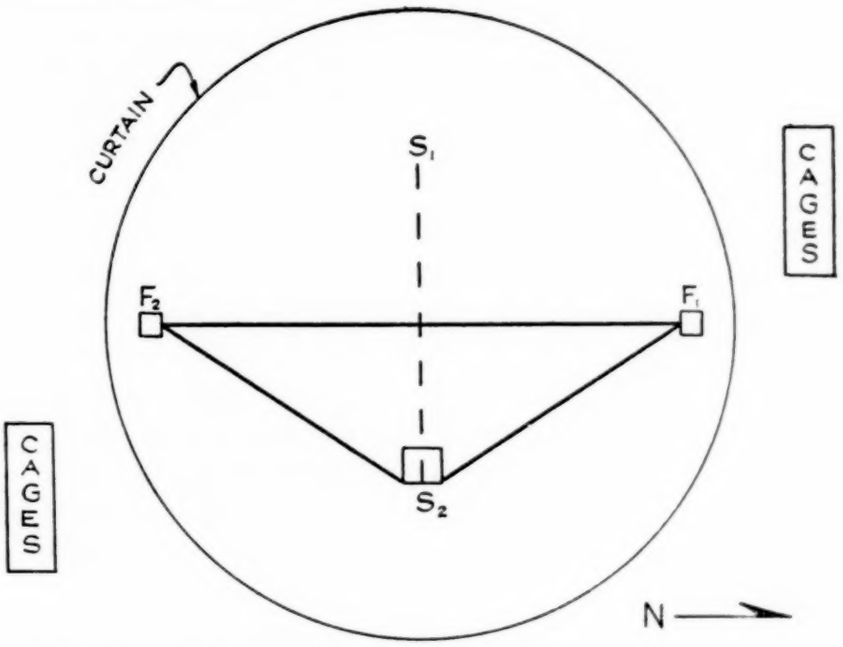

Fig. 2. Floor plan of the maze used in Experiment II. A typical short-cut setup. The dotted lines indicate the straight starting pathways which were removed when the short-cut tests were given.

The short-cuts which were added to this apparatus on the critical trials at the termination of the experimental training were two-inch wide paths similar to pathway F-1 F-2. They were attached to a platform eight in. square which was placed at one of the starting positions and extended in opposite directions to each of the two food platforms. The regular starting pathways were removed for the critical test trials. The apparatus was placed inside the circular curtain in the same position occupied by the maze in the previous experiment. The home cages of the animals, which were the only ones in the laboratory at the time, were placed symmetrically about the apparatus.

D. Procedure

Five days before the experimental training the animals were placed on a 24-hour maintenance schedule. On the two days preceding the experimental training the animals were given preliminary training on pathway F-1 F-2, which consisted in allowing each animal to eat for five min. on each of the two food platforms. On the final night of preliminary training the animals were also allowed to run down the pathway to food five times.

In the experimental training the animals were trained always to go either to the right or to the left at the choice point. A preliminary trial with no food on either platform was given each animal to determine his initial turning preference. On all subsequent trials he was trained against this initial preference. As a result of this test eight animals were trained always to go to the right and two were trained always to go to the left.

Each animal was given four trials a day with starting positions rotated in an ABBA order. If an animal was being trained to go always to the right and was started from S-1 he found food at F-2, and conversely, if started from S-2, he found food at F-1. The procedure for an

[3] The small black curtains in front of the food platforms were removed before this experiment was begun since the visual acuity of the white rat was not considered high enough to enable it to see a small pellet of food four feet away.

animal being trained to go always to the left was the converse. This training was continued for 12 days.

As an animal met the criterion of 10 successive errorless runs the straight starting pathways were removed and the short-cut pathways put in on the side of the next regular starting position. Food was placed on both platforms on all the short-cut test runs. The animal was then brought from his home cage and placed on the eight-in.-square platform facing pathway F-1 F-2 and released. On the following day he was given one trial on the double-T maze with food only at the correct platform followed by another test on the short-cut with food at both places. This procedure was followed until the experiment was discontinued on the 12th day.

E. Results

Three of the 10 animals failed to learn within 48 trials to go always to the right (or left) to obtain food. Instead they developed stereotyped habits of turning always in the wrong direction at the choice point, running to the empty food platform, and then to the correct one. The remaining seven rats met the criterion in an average of 11.1 trials. When these animals were tested on the short-cut immediately after meeting the criterion, four took the short-cut to the correct place, i.e., the one to which they had been trained to go from the given starting position, and three took the short-cut to the incorrect place.

The results obtained on the following days when the animals were given both a reinforcing trial on the double-T and a test trial with the short-cuts in place indicated that the animals made no improvement in going to the correct place over the short-cuts, and that individual rats were inconsistent in their choices during the first few days of the tests. As the tests continued from day to day the animals became consistent in their choices of short-cut, but these choices seemed to bear no relation to the previous training.

It is concluded that the animals showed no spatial-orientational ability in this situation. It should be noted that this fact cannot be taken as proof that the rat is incapable of acquiring a spatial orientation, even in the precise sense in which it has been defined, but only that our rats in our situation did not exhibit it.

Summary

1. Spatial orientation is defined in the present paper as that condition within an organism which enables the organism to make a response to a spatial relationship of objects in the environment, which relationship is determined solely by the relative positions of the objects in the environment and the organism itself. The two experiments reported here were performed in an attempt to determine whether the white rat is capable of such spatial orientation.

2. In Experiment I 32 rats were trained to go always to the left on a modified double-T elevated maze with curved starting pathways.

This maze was placed in an environment which was homogeneous with regard to visual and auditory extra-maze cues. Twenty-six animals learned the criterion (10 successive errorless runs) within 48 trials and were each given one critical trial with a short-cut added. The Experimental Group ($N = 13$) were tested with a short-cut leading to the correct place. The Control Group ($N = 13$) were tested with a short-cut leading to the incorrect place. Nearly all of the animals in both groups took the short-cut regardless of whether it led to the correct place or the incorrect place, and in spite of the fact that taking it involved a right turn never before rewarded. This result is interpreted as equivocal, neither affirming nor denying the rat's ability to gain a spatial orientation.

3. In Experiment II 10 rats were trained to go always to the right (or left) on a conventional double-T elevated maze placed in the same homogeneous environment as that of Experiment I. Seven of the rats learned within 48 trials. Each of these rats was then presented two short-cuts simultaneously, one leading to the correct place and one to the incorrect place. No preference was exhibited for the correct short-cut, even though the tests were continued for several days after the animals first met the criterion.

4. It is concluded that the negative results reported here do not prove that the rat is incapable of acquiring a spatial orientation, but only that the white rats used in these experiments exhibited no evidence of having acquired a spatial orientation in the sense in which the term was defined.

(Manuscript received August 4, 1947)

References

1. Dashiell, J. F. Direction orientation in maze running by the white rat. *Comp. Psychol. Monogr.*, 1930, **7**, No. 32.
2. Ritchie, B. F. Studies in spatial learning. III. Two paths to the same location and two paths to two different locations. *J. exp. Psychol.*, 1947, **37**, 25–38.
3. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning. I. Orientation and the short-cut. *J. exp. Psychol.*, 1946. **36**, 13–24.
4. Tolman, E. C., Ritchie, B. F., & Kalish, D, Studies in spatial learning. II. Response learning versus place learning. *J. exp. Psychol.*, 1946, **36**, 221–229.
5. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning. IV. The transfer of place learning to other starting paths. *J. exp. Psychol.*, 1947, **37**, 39–47.

THE INFLUENCE OF INTENSITY OF UNCONDITIONED STIMULUS UPON ACQUISITION OF A CONDITIONED RESPONSE[1]

BY GEORGE E. PASSEY

Tulane University

Introduction

The purpose of this experiment is to investigate the relationship between the intensity of the unconditioned stimulus and the rate of acquisition of a conditioned response. The rôle of stimulus intensity in the production of the conditioned response was first studied by Pavlov (9). He performed a series of systematic studies dealing with the effect of stimulus intensity upon size of the conditioned response and the ease with which it could be extinguished. Pavlov found that the conditioned stimulus must be of a certain magnitude before it is effective at all. Above this intensity an increase in strength serves to increase the size of the resulting conditioned responses and to prolong the number of unreinforced presentations necessary to extinguish the response. Pavlov also found an upper limit of conditioned stimulus intensity beyond which there is a decrement in the size of the response and a decrease in the number of unreinforced trials necessary for extinction.

The effect produced by a variation in intensity of the conditioned stimulus was altered by a change in the intensity of the unconditioned stimulus. A relatively strong stimulus could be made into a conditioned stimulus despite its unsuitability by increasing the strength of the unconditioned stimulus. This observation led Pavlov to the conclusion that the conditioned stimulus and its normal response must be weaker, and biologically less important than the unconditioned stimulus and its response. Further lending support to the notion of stronger conditioned stimuli was the discovery of the principle of summation. The conditioned response elicited by the simultaneous presentation of two weak conditioned stimuli was larger than that produced by the application of one stimulus alone.

Pavlov was not concerned with the rate of acquisition of the response as a function of stimulus intensity. Carter's experiment (3) represents a

[1] This paper represents the main content of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in the Division of Psychology at Clark University. The author is indebted to Dr. Robert H. Brown for his direction of the research.

systematic study of the effect of the intensity of the conditioned stimulus upon the rate of acquisition of a conditioned eyelid response. He used a puff of air to the cornea as an unconditioned stimulus to elicit a wink and conditioned this response to a tonal stimulus which was varied throughout six different intensities separated by equal intervals of 15 db. The rate of conditioning as measured by the number of anticipatory responses during a series of 50 reinforced presentations of the conditioned stimulus was not affected by the intensity of the conditioned stimulus.

Weber and Wendt (11) investigated the acquisition of the conditioned eyelid response in humans under varying conditions of reinforcement. A light was used as the conditioned stimulus. The four conditions of reinforcement were: (a) sound, (b) either a puff of air or a sound used in random order, (c) a puff of air to the cornea, and (d) a puff of air to the cornea plus sound. Sound when used alone was the least effective reinforcement. The combination of the sound with the puff of air was the most effective reinforcement. The use of air alone and presentation of a puff of air or sound in random order were about equal in effectiveness.

Humphreys (8) studied the extinction of conditioned responses following two conditions of reinforcement in a study closely related to the present investigation. Conditioned psychogalvanic responses were established in an experimental group of 40 *Ss*. Twenty *Ss* received reinforcement in all conditioning trials; a second group of 20 *Ss* was given reinforcement on only half of the conditioning trials. Varying conditions of reinforcement were not effective in altering the results.

The effect of stronger conditioned stimuli has been shown to increase the size of resulting conditioned responses and to strengthen their resistance to extinction. The use of stronger unconditioned stimuli accomplished by the use of two stimuli as opposed to one stimulus alone has been shown to be more effective in conditioning. The rate of conditioning has been demonstrated to be unaffected by the variation in intensity of the conditioned stimulus. The variation in strength of the unconditioned stimulus and its effect upon the rate of conditioning have not been studied, and Hilgard and Marquis (5) have indicated a need for exploration of this variable. The present investigation, therefore, undertakes such a study. The eyelid responses of 40 *Ss* were conditioned to a tonal stimulus which was held constant for all groups. A puff of air to the cornea served as the unconditioned stimulus and was varied so that the effect of its strength upon conditioning might be observed.

Procedure

Each of four groups of 10 *Ss* was conditioned at a different intensity of air puff, while the tonal stimulus remained constant. The rate of conditioning was determined from a record of the anticipatory responses made by the *Ss* during 50 reinforced presentations of the tone. The strength of conditioning was also measured by recording the size of the conditioned responses and the number of responses during an extinction series.

The intensities of the unconditioned stimulus were selected by giving 15 *Ss* puffs at six air pressures between 2.5 and 100 lbs. per sq. in. Each *S* received five puffs at each intensity arranged in random order. We measured the sizes of the responses in mm. and selected four pressures which had produced mean responses differing in size by 10 mm. Pressures of 7.5, 18, 44, and 88 lbs. per sq. in. were selected.

The 40 *Ss* were all students at Clark University. Care was taken to see that no *S* had received instruction as to the nature of the conditioned response. *Ss* were told only that the *E* was making measurements of eye potentials and that the experiment would require about 30 min. of their time.

In determining the intensity of the conditioned stimulus for each *S*, ascending and descending thresholds were determined for a tone of 500 c.p.s. and the conditioned stimulus for each *S* was set at 30 db. above his threshold.

The experiment involved the following steps in stimulation:

1. The response to presentation of tone of intensity 30 db. above the *S*'s threshold was observed. Five trials were given with this tone alone. The duration of the tone was 450 ms. Its frequency was 500 c.p.s. All presentations were separated by intervals of 16 to 28 sec., randomly arranged by reference to the table of random numbers.

2. The *S*'s response to a presentation of air at the pressure to be used with the group of which he was a member was observed for three trials. The duration of the blast was approximately four ms. and the successive presentations were separated by intervals of from 16 to 28 sec.

3. The *S* was then tested by two presentations of the tone as in (1). All *Ss* who gave a wink to either tone were eliminated from experimentation.

4. Fifty paired presentations of tone and air blast were then given. The successive presentations were separated at random by intervals of 16 to 28 sec. The tonal stimulus had a duration as before of 450 ms. The air blast was administered for a duration of four ms. and overlapped the tonal stimulus during its terminal four ms.

5. Following the 50 paired presentations of tone and air there was an extinction series of 10 presentations of tone alone. Conditions except for absence of air puff remained the same.

6. At the conclusion of the extinction series the *S* was informed that the experiment was ended, and a short informal interview was held in which the *E* attempted to discover whether or not any uncontrolled cues had operated to influence the *S*'s behavior. A short explanation of the experiment was given.

Apparatus

The tonal stimulus was produced by a General Radio Beat Frequency Oscillator number 613B which was set for maximum output at a frequency of 500 c.p.s. The unconditioned stimulus was produced by a standard oxygen cylinder equipped with a reducing valve capable of adjustment for pressures between 2.5 and 100 lb. per sq. in.

A synchronous motor equipped with a rotating arm and metal contact arcs timed and accomplished the presentation of the stimuli through relays. Recording of the onset and duration of the stimuli was accomplished by use of signal markers excited through the relays.

The recording of the signal markers and the recording of the *S*'s responses were made on a continuous roll of paper passing through a continuous feed kymograph. The pen for the recording of the responses of the *S* was activated by a nylon thread from the *S*'s eyelid to a heart lever.

The relays and the valve concerned with the presentation of the air puff were enclosed in sound proof boxes and were kept in constant operation in order to eliminate cues. The lighting was such as to keep the *Ss*' eyelids from drooping and to render cues impossible.

Results

The number of anticipatory responses during 50 reinforced presentations and the number of conditioned responses during the 10 presentations of tone alone in the extinction series are summarized in Table I. The magnitude of the responses is also presented in the table. The data have been grouped in terms of average number and size for units of five trials. The size is expressed in mm.

TABLE I

NUMBER AND SIZE OF RESPONSES DURING 50 REINFORCED AND 10 EXTINCTION TRIALS AVERAGED FOR GROUPS OF FIVE TRIALS

Trials	Intensity of air puff in lb. per sq. in.							
	7.5		18		44		88	
	No.	Size (mm.)	No.	Size (mm.)	No.	Size (mm.)	No.	Size (mm.)
1–5	.4	1.7	1.3	5.1	1.3	4.3	1.0	4.3
6–10	1.7	6.7	3.0	12.4	3.3	13.2	3.6	21.0
11–15	2.2	7.6	3.6	14.1	3.2	14.1	4.2	17.9
16–20	2.2	7.9	3.5	13.0	3.6	15.6	4.5	17.5
21–25	3.0	9.7	3.8	13.9	4.2	15.6	4.7	19.3
26–30	3.1	10.0	3.9	14.6	4.3	16.3	4.9	18.5
31–35	3.3	9.5	4.3	16.3	4.8	17.7	5.0	20.2
36–40	3.3	10.9	4.0	15.4	4.3	17.3	4.7	17.5
41–45	3.1	10.4	4.5	16.6	4.5	17.1	4.8	18.9
46–50	3.5	10.2	4.1	15.5	4.6	16.3	5.0	19.9
Extinction Series								
1–5	1.7	3.4	2.6	5.3	3.1	6.0	2.7	8.6
6–10	.5	.3	1.7	2.3	1.5	2.7	.5	.8

Fig. 1 illustrates the results presented in Table I. The air pressures for the unconditioned stimulus are as indicated. It is clear that the mean number of responses for each group of five trials increases as the total number of reinforced presentations increases,

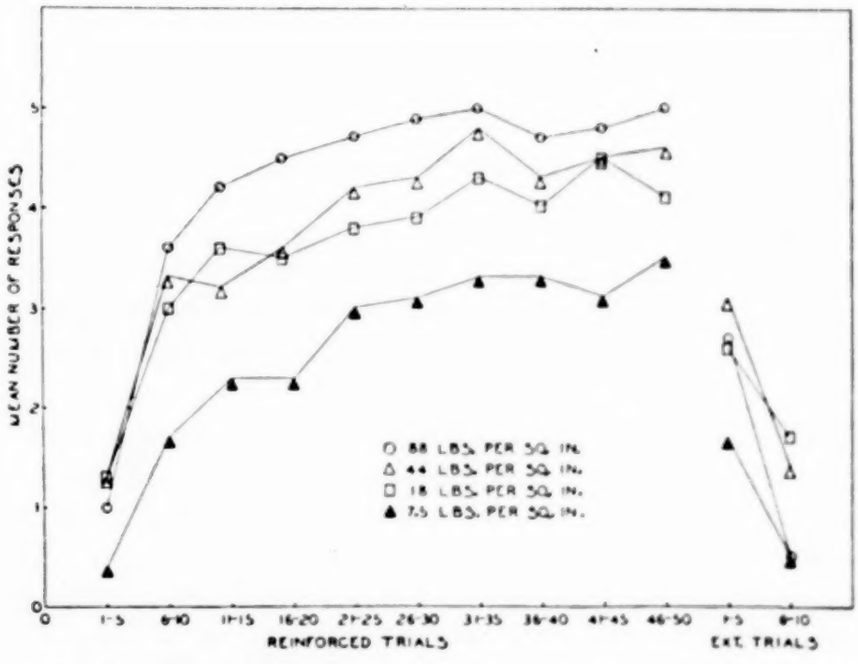

FIG. 1. Mean number of anticipatory responses for each group of five trials during 50 reinforced and 10 unreinforced trials

and that the stronger puffs of air have an advantage in the production of a faster rate of conditioning. The plot of the conditioned responses during the extinction series does not indicate as definite a relationship between the number of responses and the intensity of the unconditioned stimulus. The curve of acquisition of the conditioned eyelid response is a negatively accelerated one, and the shape of this curve apparently does not change with an increase in the intensity of the unconditioned stimulus, but merely shifts on the ordinate.

The data given in Table I for the mean size, in mm., of the anticipatory responses and for the extinction responses are plotted in Fig. 2.

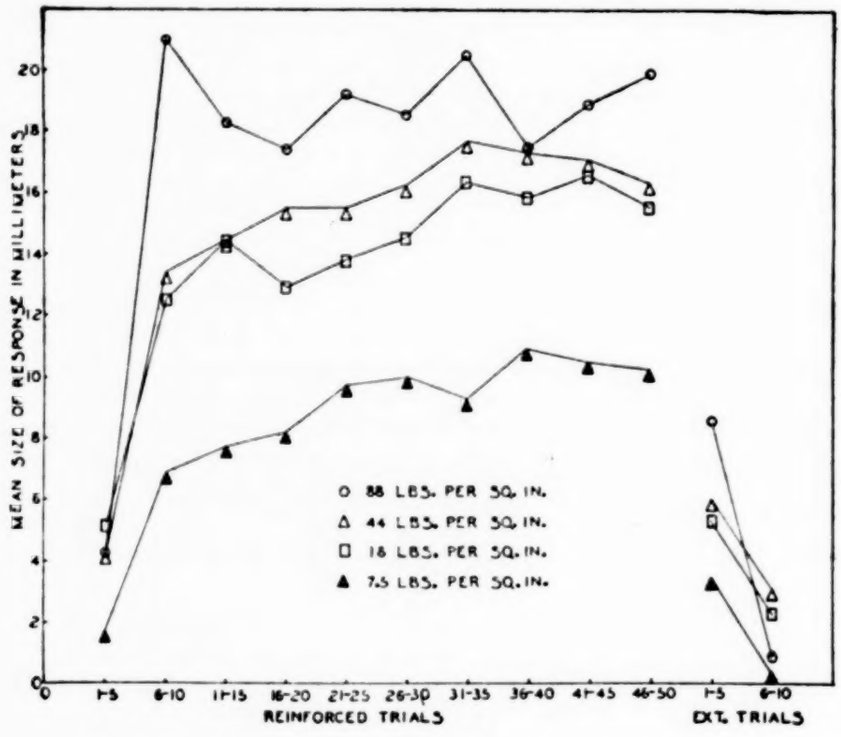

FIG. 2. Mean sizes of anticipatory responses for groups of five trials during 50 reinforced and 10 unreinforced trials

These responses are averaged for each group of five trials and the means are plotted. The advantage of stronger intensities of the unconditioned stimulus in producing large conditioned responses is evident.

In Table II the number of anticipatory responses made during 50 reinforced trials for each *S* is indicated under a column headed by the intensity of unconditioned stimulus used in conditioning that *S*. A plot of the data of Table II is presented in Fig. 3. The mean number of responses for groups of 10 *S*s conditioned at a single intensity of the unconditioned stimulus is plotted as a function of the logarithm of the air pressure of the unconditioned stimulus. The rate of conditioning as measured by the number of anticipatory responses

TABLE II

NUMBER OF ANTICIPATORY RESPONSES IN 50 REINFORCED TRIALS FOR FOUR INTENSITIES OF UNCONDITIONED STIMULUS

Subject	Unconditioned stimulus intensity indicated in lb. per sq. in.			
	7.5	18	44	88
1	33	41	22	46
2	34	45	44	40
3	0	47	42	46
4	44	44	18	43
5	36	0	43	41
6	29	48	46	44
7	33	36	45	41
8	44	23	46	37
9	5	42	28	42
10	0	34	46	43

in 50 reinforced presentations varies directly with the logarithm of the air pressure used as the unconditioned stimulus.

In determining the significance of the obtained results the method of analysis of variance was used. An F of 4.38 was necessary for significance at the one percent level. The obtained F was 3.066 and was significant between the five and one percent levels.

In order to find the level of significance between pairs of groups a computation of t was made. The t's for differences between the groups are indicated in Table III. The difference between the group

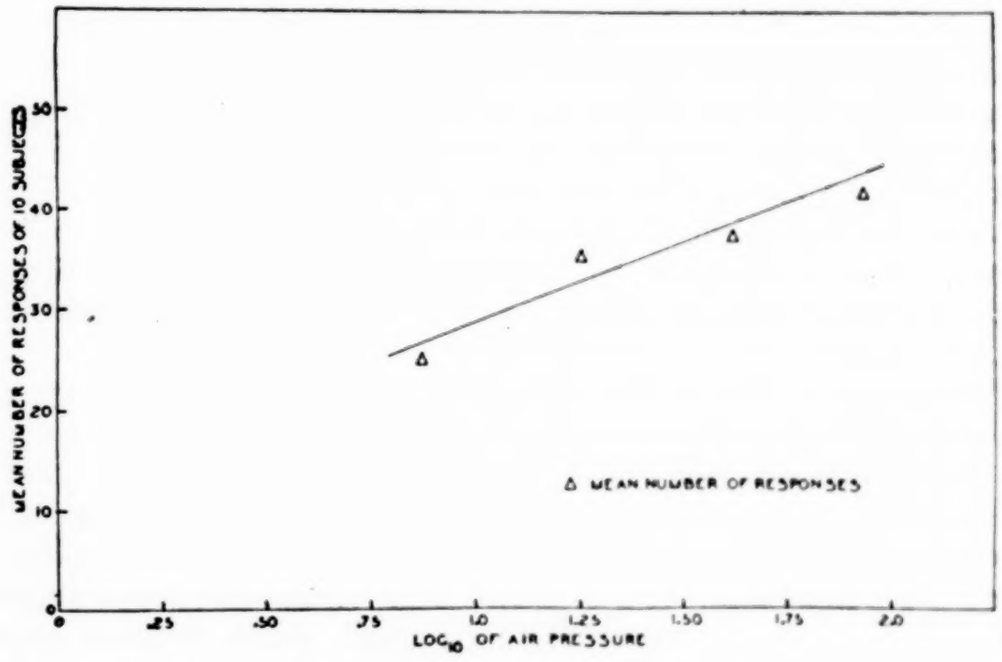

FIG. 3. The relationship between the average number of anticipatory responses of 10 *Ss* at each stimulus intensity and the logarithm$_{10}$ of the strength of air puff used as unconditioned stimulus.

TABLE III

VALUES OF *t* FOR DIFFERENCES BETWEEN PAIRS OF INTENSITIES OF UNCONDITIONED STIMULI

7.5 and 18 lb. per sq. in.	1.76*
7.5 and 44 lb. per sq. in.	2.14
7.5 and 88 lb. per sq. in.	2.89
18 and 44 lb. per sq. in.	.35
18 and 88 lb. per sq. in.	1.10
44 and 88 lb. per sq. in.	.75

* At 10 percent level, $t = 1.64$; at one percent level, $t = 2.58$.

conditioned at 7.5 lb. per sq. in. and the group conditioned at 88 lb. per sq. in. is highly significant.

A product-moment coefficient of correlation between the number of conditioned responses during the 50 reinforced trials and the number of responses during the 10 extinction trials is $+ .57 \pm .07$. A calculation of *eta* shows that the deviation of the regression from linearity is highly significant and that the product-moment *r* is not a satisfactory measure of the relationship.

DISCUSSION

The general course of acquisition of the conditioned eyelid response is similar to the familiar negatively accelerated learning curve. Stimulus intensity of the air puff exhibits its effect by raising the number of anticipatory responses in each group of five trials. The rate of acquisition of the conditioned response, as shown by the larger number of anticipatory responses during the 50 reinforced trials, can be observed to operate throughout the course of conditioning. The number of anticipatory responses made during the 50 reinforced trials varies directly with the logarithm of the air pressure of the stimulus. In determining the intensities of air puff to use as stimuli in this experiment, we found that the size of the response to the air puff varied as the logarithm of the air pressure used in stimulation. Indirectly it can be seen that the rate of acquisition of the conditioned eyelid response is proportional to the magnitude of the unconditioned eyelid response.

The curves of Fig. 1 also uphold the postulates of Hull (7) and tend to substantiate his amount-of-reinforcement hypothesis. The limit of practice was not approached in this investigation, yet, from an inspection of the curves, it would seem that increased intensity of the reinforcing puff has given a higher final level of potential habit strength as well as a larger increment with each succeeding reinforcement with the higher intensities of air puff. More positive results would have been obtained if the limit of practice had been approached.

The size of the conditioned response is also a function of the intensity of the unconditioned stimulus. The curves of Fig. 2 are of the

general form of those found by Hovland (**6**). The magnitude of the conditioned galvanic skin response, in his experiment, increased rapidly during the first stages of conditioning and the levelled off and remained practically constant during the latter half of the conditioning procedure. Gantt (**4**) has shown that response size is a function of the intensity of the unconditioned stimulus and our results agree with his. Other experimenters have shown that the size of the conditioned response is a function of the intensity of the conditioned stimulus and the present finding supplements their work.

The analysis of the data yielded a highly significant difference between the groups conditioned at highest and lowest intensity of air puff. The differences in performance of the groups conditioned at air pressures more nearly alike in intensity gave little evidence of significance. It is possible that the range of intensities should have been broadened to include more widely divergent pressures. Had this been done more differences approaching significance might have been found.

An interesting finding was that of the curvilinearity of the regression when a Pearson *r* was computed between the number of anticipatory responses during 50 reinforced trials and those in the 10 trial extinction series. Our correlation coefficient may be compared with those obtained by other experimenters using the eyelid response. Bernstein (1) and Campbell (**2**) have reported coefficients of correlation for this response. Bernstein reported a rho of $+ .88 \pm .02$ between the number of anticipatory responses during 50 reinforced trials and the number of trials necessary to obtain extinction. Campbell, whose procedure more nearly follows that of the present investigation, reports an *r* of $+ .66 \pm .06$ which agrees rather well with our *r* of $0.57 \pm .07$. He calculated the *r* between the number of anticipatory responses during 50 reinforced trials and the number of responses during an extinction series of 10 trials. These experimenters give no indication as to whether a check of linearity of regression was made.

According to Skinner (**10**), who postulated the concept of reflex reserve as a measure of the strength of conditioning in contradistinction to the size of response, a curvilinear relationship would be expected. Skinner describes conditioning as essentially the building up of this reflex reserve. The amount of reserve determines the rate of extinction. The reserve builds up rapidly at first and its rate of increase drops off as the reinforcements are continued. This would account for the curvilinear relationship we found.

It is concluded that the rate of acquisition of a conditioned eyelid response is a function of the logarithm of the stimulus intensity of the unconditioned stimulus, and, further, that the size of the resulting

conditioned responses is a function of the intensity of the unconditioned stimulus.

Summary

The purpose of this experiment was to determine the effect of four different intensities of unconditioned stimulus upon the rate of acquisition of the conditioned eyelid response.

Forty students at Clark University served as *S*s. Ten *S*s composed an experimental group, and each group of 10 *S*s was conditioned to wink to a tonal stimulus reinforced by a puff of air to the cornea. All members of a given experimental group were conditioned to the same intensity of air puff. Air puffs of 7.5, 18, 44, and 88 lb. per sq. in. were used, and the tonal stimulus was held constant at 500 c.p.s. at an intensity of 30 db. above the threshold of the subject.

Analysis of the results indicates that the rate of acquisition of the conditioned eyelid response is a function of the intensity of the unconditioned stimulus. The rate of acquisition bears a direct relationship to the logarithm of the intensity of the unconditioned stimulus. The size of resulting conditioned responses is a function of the intensity of the unconditioned stimulus.

(Manuscript received August 11, 1947)

References

1. Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.
2. Campbell, A. A. The interrelations of two measures of conditioning in man. *J. exp. Psychol.*, 1938, **22**, 225–243.
3. Carter, L. F. The intensity of the conditioned stimulus and rate of conditioning. *J. exp. Psychol.*, 1941, **28**, 481–490.
4. Gantt, W. H. The nervous secretion of saliva: the relation of the condition reflex to the intensity of the unconditioned stimulus. *Amer. J. Physiol.*, 1938, **123**, 74.
5. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton Century, 1940.
6. Hovland, C. I. The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1936, **21**, 261–276.
7. Hull, C. L. *Principles of behavior.* New York: Appleton Century, 1943.
8. Humphreys, L. G. Extinction of conditioned psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–75.
9. Pavlov, I. P. *Conditioned reflexes.* (Trans. G. V. Anrep.) London: Oxford Univ. Press, 1927.
10. Skinner, B. F. *Behavior of organisms.* New York: Appleton Century, 1938.
11. Weber, H., & Wendt, G. R. Conditioning eyelid closure with various conditions of reinforcement. *J. exp. Psychol.*, 1942, **30**, 114–124.

'SPONTANEOUS RECOVERY' OF VERBAL ASSOCIATIONS

BY BENTON J. UNDERWOOD

Northwestern University[*]

I. Introduction

Data from a recent study (4) suggested that under certain conditions a process occurs with verbal associations which opposes the forgetting usually associated with the passage of time. Because the conditions which produce such a process may be highly specialized, this recent study will be summarized briefly.

Ss learned two lists of paired adjectives to a criterion of one perfect trial. Both lists had the same stimulus words but different responses. Two such lists were learned in each of four conditions, the recall and relearning for the four conditions being: (1) first list after five hr., (2) second list after five hr., (3) first list after 48 hr., and (4) second list after 48 hr. Recalling the first list follows the retroactive inhibition paradigm, while recalling the second list follows the operations for proactive inhibition. The study showed that after five hr. the recall of the first list was significantly less than the recall of the second list, while after 48 hr. there was no difference in the recall of the two lists. Furthermore, the recall and relearning scores of the first list were as high after 48 hr. as after five.

The recall findings after 48 hr. were shown to be in conflict with an implication of the unlearning theory of retroactive inhibition. This theory states that during the learning of a second list the first list responses are unlearned or weakened as a consequence of their nonreinforcement. Such a process should leave the second of the two lists relatively stronger. In the experiment under consideration this implication was not substantiated in the 48-hr. recall measures. This finding, plus the finding that the recall of the first list was as good after 48 hr. as after five, led to the following proposals:

1. Unlearning is to be conceived of as analogous to the extinction of a conditioned response, and

2. Like extinguished conditioned responses, unlearned verbal associations recover some of their response-strength with the passage of time. It is this process which is described as running counter to the usual forgetting which takes place with the passage of time.

[*] This study was made possible by a grant from the Ethel Dummer Fund, University College, Northwestern University.

The present study was performed to explore further this supposed recovery tendency. An attempt has been made to measure the strength of response tendencies without the presence of complicating factors which may obtain when *S* is requested to recall a specific list. To accomplish this aim a modified free recall technique has been used in which *S* is presented with the stimulus word common to both lists which have been learned, with the instructions to give the first of the two responses which come to mind. It was thought that this method would eliminate failure of response due to (1) *S*'s 'fear' of making an error, (2) inadequate set at the time of recall, and (3) the limited recall time allowed by the two-sec. anticipation interval.

II. Procedure

Subjects and materials.—The volunteer *S*s for this experiment were 12 men and 12 women. The learning material, lists of 10 paired two-syllable adjectives, was exactly the same as used in the previous study. These lists were of the A–B, A–C relationship in which the second list had the same stimulus words as the first list but different responses. A Hull-type drum was used to present the material. To prevent serial learning, the words were presented in three different orders within a list. A two-sec. anticipation period was used and eight sec. rest followed each trial.

Procedure.—The learning and recall procedures for the four conditions were exactly the same, the conditions differing only in the length of the interval elapsing between learning and the recall test. For each condition *S* learned two lists in immediate succession, each to a criterion of one perfect trial. The four conditions then were: recall after one min., after five hr., after 24 hr., and after 48 hr.

Four sets of experimental lists were employed with Set I being always used on the first experimental day, Set II on the second experimental day, etc. The four experimental conditions were completely counterbalanced for the 24 *S*s, thus equalizing for all conditions the differences in scores arising from practice effects and lists of unequal difficulty. Two practice days preceded the four experimental days. On each of these practice days *S* learned two lists and was introduced to the modified free recall method.

In conducting the free recall the common stimulus words for the two lists were presented on the same kind of heavy white tape used for presenting the words during learning. A push-button switch was placed in circuit with the power for the drum motor so that the drum could be conveniently started and stopped by *E*. After learning the two lists on the first practice day, the following instructions were read to *S*:

> Now I am going to present the first members of the pairs of the two lists which you learned. You will remember, of course, that in memorizing these lists you have learned two responses to each stimulus word. The stimulus words will be given now without the response words. As soon as a stimulus word comes into the window I want you to give the *first* of the two possible responses which may come to your mind. It is absolutely essential that you give the first of these words which occurs to you.

E then returned to his chair, prepared himself for the recall and then said:

> We are ready now. Remember to be sure and say the first of the two response words which might come to you. Give only one response. The closing of the shutter will let you know that we are moving on to another word.

These instructions were repeated again on the second practice day preceding the recall of the lists learned on that day. However, on the four experimental days, the instructions were briefed as follows:

> As before, I am going to present the stimulus words of the last two lists learned and I want you to give the first of the two response words which might occur to you.

In starting the free recall *E* brought the first stimulus word into the window, at the same time starting a stop watch. If no response had been given by the end of 10 sec., *E* asked, "What response *does* this stimulus word make you think of?" If no response was given by the end of 20 sec., *E* instructed: "Give any word which occurs to you." This latter instruction was repeated every 10 sec. until a response was given or until one min. had passed, whichever came first. Practically all responses from the lists, if given at all, were given during the first few seconds. Responses given after 10 sec. were for the most part from outside the specific lists. In a few cases no response of any kind was obtained. In each case *E* recorded the time between the presentation of the stimulus word and the giving of the response.

After a response to the first stimulus word was obtained, *E* moved the drum on to the second stimulus word and so on down through the list. No relearning trials were given and *S* never saw the responses from the two lists during the free recall. During learning three different orders of presentation of the pairs were used. On the recall trial this order of presentation was maintained. Thus, if the last learning trial of the second list had been under order No. 2, order No. 3 was used on the recall trial.

The stimulus words for the free recall were always presented on the left side of the drum. Each condition occurred 24 times. For 12 of these occurrences the first list was learned on the left side of the drum and the second on the right, while for the other 12 occurrences the locations were reversed. The results should not then be biased by having either the first or second list always learned on the left side of the drum—the same side on which the stimulus words for the free recall were presented.

III. Results

Original learning.—The number of trials required to reach one perfect recitation for the two lists for the four conditions is shown in Table I. These scores and error measurements are very much the

TABLE I

Mean Number of Trials Required to Reach One Perfect Trial for Both Lists for All Conditions

	1 min.		5 hr.		24 hr.		48 hr.	
	M	σ_M	M	σ_M	M	σ_M	M	σ_M
First List	11.17	.86	11.88	1.11	12.33	1.07	12.13	1.09
Second List	10.83	.97	10.92	1.16	10.25	.83	9.96	.76

same as those reported for the original learning in the previous experiment with another group of *S*s. The expected positive transfer is evident in terms of the tendency for all lists learned second to be learned in fewer number of trials than the first lists. As can be seen from Table I, there is little difference in the mean number of trials required to learn the first list under the various conditions and likewise little difference among the second lists.

Recall results.—The responses evoked on the free recall trials are analyzed in Table II. Reliability measures are given whenever the frequency of response justifies them.

It is to be noted first that the frequency of response from the first list shows no over-all decrease from one min. to 48 hr. There is a slight increase in the frequency of response between one min. and

five hr., and a mean increase of .46 items between five and 24 hr., after which there is a decrease. None of these increases in the frequency of response from the first list is statistically significant. However, the facts that no forgetting of the first list occurs over a 48-hr. period and that the means suggest an increase in the frequency of first list responses are submitted as evidence for inferring a process which runs counter to factors which normally bring about a decrease in response strength of verbal associations with the passage of time. This process may be similar to that which brings about spontaneous recovery of extinguished conditioned responses.

The frequency of the responses from the second list shows a marked decrease from one min. to 48 hr. This would be expected according to the usual 'laws' of forgetting. The frequency of responses from this second list at 24 hr. is somewhat greater than would be expected on the basis of a smooth curve drawn with regard to the

TABLE II

MEAN NUMBER OF RESPONSES OF FIVE DIFFERENT CATEGORIES WHICH OCCURRED ON THE FREE RECALL TRIALS

	1 min.		5 hr.		24 hr.		48 hr.	
	M	σ_M	M	σ_M	M	σ_M	M	σ_M
1. Response from 1st List	2.42	.29	2.71	.40	3.17	.24	2.50	.27
2. Response from 2nd List	7.13	.27	5.13	.47	4.79	.35	3.33	.25
3. Word from Previous Lists	.04	—	.17	—	.63	—	1.25	.31
4. Word from Outside the Lists	.04	—	1.29	—	.71	—	1.92	.38
5. Miscellaneous and No Response	.37	—	.71	—	.71	—	1.00	—

frequency for the other three intervals. It is clear, however, that the functional properties of the first list responses are considerably different from the properties of the second list responses. The second list, having not been unlearned or extinguished, shows the usual forgetting; the first list, having been unlearned, does not exhibit usual forgetting.

Fig. 1 shows the basic results in graphic form. In plotting Fig. 1 all responses other than those from the first and second lists are added together to give the points for the 'other' curve.

The increase in the frequency of responses from the first list with the passage of time relative to the decrease in frequency of the second list responses may be demonstrated in another manner. Using the total responses from both lists as a base, we may calculate the percent response from the first list for the various intervals. Thus, after one min., a mean of 9.55 responses occurred all of which were 'correct' responses from the first and second lists. Of these, 2.52 or 25.3 per-

cent were from the first list. For the remaining three intervals in order the percents are 34.6, 39.8, and 42.9. Relative to the frequency from the second list, then, the frequency of responses from the first list increases throughout the 48 hr.

The components of the 'other' curve in Fig. 1 require further analysis. Category 5 in Table II (miscellaneous) constitutes a relatively low frequency of the total response picture. Category 3,

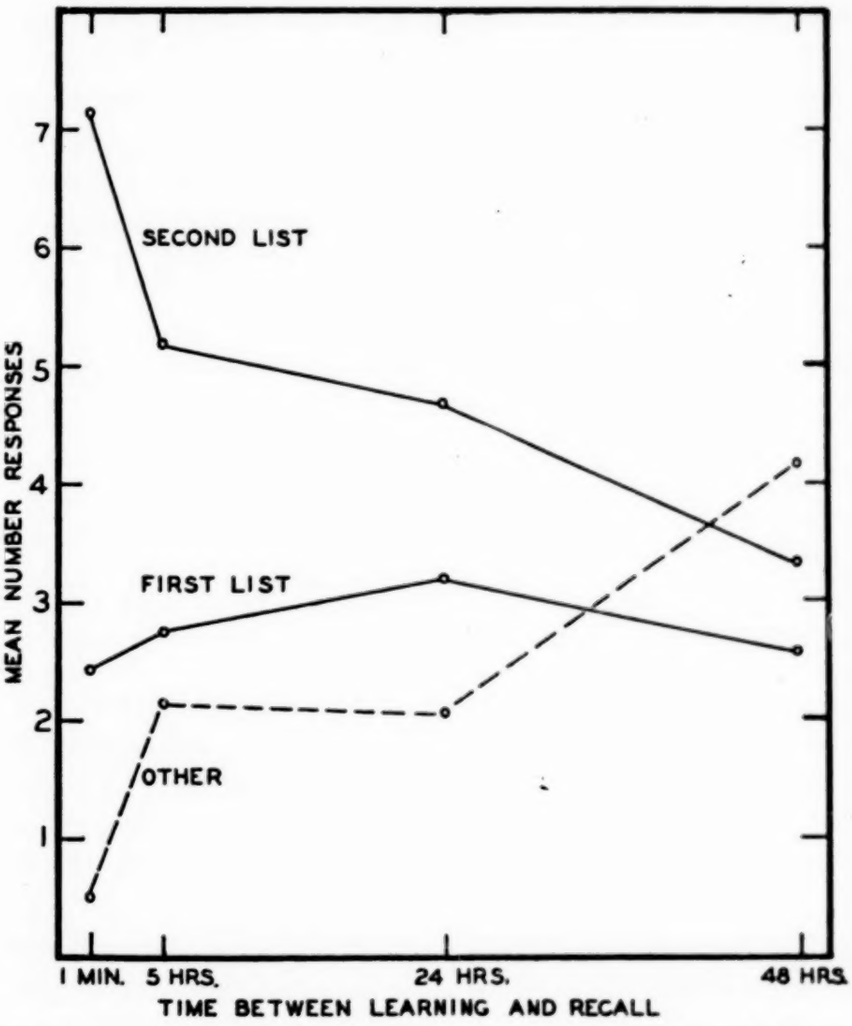

FIG. 1. Mean frequency of response on the free recall trial for the four different time intervals. The curves show the frequency from the first list, from the second list, and a single curve for all 'other' responses.

words from previous lists, shows a progressive increase with the passage of time and by 48 hr. makes up 12.5 percent of the total responses. Words which were given as responses and which occurred in no lists formally learned constituted nearly 20 percent of the total responses given at 48 hr. An attempt has been made to analyze further these two response categories (3 and 4 in Table II) to discover if they held any systematic relationship to the 'correct' stimulus-

response pairings. It is conceivable that many of these responses came about as a result of generalization. This analysis is shown in Table III where four categories are used in the breakdown.

Stimulus generalization was assumed as the basis for a response if the response elicited was one which had been previously learned to a similar stimulus, the similarity being either formal or meaningful. Only six cases which appeared to meet this requirement were found. An example (which occurred twice) is *S* giving the response *anxious* to the stimulus *worthless* when *anxious* had been the correct response to *wayward* in a previous list. *Wayward* and *worthless* have similarity of meaning as well as some formal similarity. Judgments concerning the occurrence of stimulus generalization could be made, of course, only for stimuli which occurred in previous lists. Many of the responses which have been classified as 'no relation' may have been due to generalization to stimuli of which *E* is unaware.

TABLE III

ANALYSIS OF RESPONSES FROM NEITHER THE FIRST NOR SECOND LISTS
See Text for Complete Definition of Categories

	1 min.	5 hr.	24 hr.	48 hr.
Stimulus Generalization	0	0	3	3
Response Generalization	0	12	8	4
Free Association	1	5	4	17
No Relation	1	18	17	52

The second category in Table III, response generalization, appeared clear-cut in most cases. There is a tendency for responses of this type to decrease in frequency with the passage of time, with the peak at five hr. Examples of response generalization are as follows: saying *resting* instead of *restful; misty* instead of *filmy; inane* instead of *witless; gaiety* instead of *festive.*

The free association category consists largely of a reflection of the stimulus. Examples are: to the stimulus *glassy* the response *slippery;* to the stimulus *pious* the response *righteous;* to the stimulus *obscene* the response *dirty;* to the stimulus *hollow* the response *empty.* In general, these free-association type responses increase with the passage of time.

It can be seen that the greatest frequency of responses which were neither from the first or second lists could not be identified as to possible causal basis. As mentioned before, many of these may be due to stimulus generalization the sources of which cannot be traced. It cannot be said, of course, merely by examining Table III, that response generalization occurs with greater frequency than stimulus generalization.

Latencies.—*E* recorded the latency of each response by a manually operated stop watch. The latencies are probably inaccurate in terms of absolute measurements since constant errors undoubtedly operated. In recognition of the gross character of the measurements, the data have been compiled into three categories: 0–5.0 sec., 5.1–10.0 sec., and 10.1 sec. or over. The frequency of response for each latency category for each condition is shown in Table IV. Since different

TABLE IV

LATENCIES OF THE RESPONSES ON THE FREE RECALL TRIAL

The data show the frequency of responses for each of the three latency categories along with the percent frequency. The table is to be read as follows: After 1 min., 58 responses from the first list were given. Of these, 51, or 88 percent, were given within 5 sec., 9 percent had latencies between 5.1 and 10.0 sec., and 3 percent had latencies longer than 10.0 sec.

	Latency (sec.)	1 min.		5 hr.		24 hr.		48 hr.	
		N	%	N	%	N	%	N	%
First List Responses	0–5.0	51	88	50	77	60	79	37	62
	5.1–10.0	5	9	10	15	8	10	14	23
	10.1 and over	2	3	5	8	8	10	9	15
Second List Responses	0–5.1	151	88	80	65	92	80	55	69
	5.1–10.0	12	7	28	23	20	17	15	19
	10.1 and over	8	5	15	12	3	3	10	12
All Other Responses	0–5.0	2	18	8	16	10	21	5	5
	5.1–10.0	7	64	9	18	13	27	22	23
	10.1 and over	2	18	33	66	25	52	67	71

categories had different frequencies, percent frequency is shown as the basic comparative measure. Comparison of the percent frequency of each latency category for items from the first or second lists ('correct' responses) shows little difference. The data show clearly, however, that the responses which were from neither the first or second list had much longer latencies than the 'correct' responses. This is significant with regard to evaluating the free recall technique, since it would appear to indicate that the instructions were successful in actually making *S* 'search' for responses from the lists before giving another response. It is highly likely that *S* knew that some of these long latency responses were not the 'correct' responses.

Evaluation of the method.—The modified free recall method was used in the present study in an effort to get an uncomplicated index of the relative response strengths of the items in the two lists. It was believed that in the usual recall procedure after fairly long intervals of time, some correct responses which were 'available' were not being given by *S*. It was conceivable that such a bias might have been greater for one of the lists than for the other, and might also

vary as a function of the length of the recall period. Since it is traditional to use specific recall with a limited anticipation interval, the present data should be evaluated against data acquired by the traditional method. The previous study (4) supplies the necessary data for the traditional form of recall at five and 48 hr. Recall data obtained by the two methods for these two intervals are shown in Table V.

As was expected, the free recall produced a greater frequency of response than did the traditional recall. Because of this, comparisons must be made on a relative basis. At five hr. both methods yielded 35 percent response from the first list and 65 percent from the second. At 48 hr. such equality does not obtain.

TABLE V

COMPARISON OF THE RECALL OBTAINED BY THE TRADITIONAL RECALL METHOD AND THE FREE RECALL METHOD AFTER FIVE AND 48 HOURS. THE FIGURES INDICATE TOTAL CORRECT RESPONSES BY 24 *Ss*

	5 hr.		48 hr.	
	1st List	2nd List	1st List	2nd List
Traditional Recall	41	75	50	51
Free Recall	65	123	60	80

After 48 hr. in the previous study, the first list produced one more response than the second, whereas by the free recall method 20 more responses were produced from the second than from the first list. Since the traditional recall method is used as the basis for evaluation, and since at 48 hr. by this traditional method the frequency of response from the two lists is divided approximately 50/50, the chi-square technique may be used to determine if the 60/80 observed frequency at 48 hr. deviates significantly from the expected 50/50 (70/70) distribution. The chi-square value is 2.86 and with one degree of freedom, P is slightly less than .10. While the 50/50 hypothesis is clearly tenable, possible differences in the results by the two methods are suggested by this divergence at the 48-hr. recall.

The free recall technique has one drawback in that no observation can be made of 'true' intrusions, i.e., items from one list intruding during the specific recall of the other. On the other hand, the forcing of responses by the free recall procedure gives a record of a considerable number of responses from neither of the lists. It has been clearly shown that the frequency of these responses increase with the passage of time. This is, no doubt, a function in part of the 'forcing' procedure, but the fact that many of these forced responses bore a

relation to the correct responses is important. Relearning scores could be obtained with the free recall method if such were desired.

IV. Discussion

The present data show that when two verbal responses are learned to the same stimulus the functional properties of those stimulus-response connections become markedly different. The association acquired second exhibits an expected decrease in effective response strength with the passage of time, while the first association shows no decrease in effective response strength with the passage of time up to at least 48 hr. From these findings it is inferred that some process which runs counter to the normal forgetting processes is operating in the case of the first learned response. The process appears to be analogous to spontaneous recovery as observed in the conditioned response following extinction.

The data further confirm the usefulness of a construct such as unlearning (**1, 2, 3**). That a weakening or unlearning or extinction of the first list responses takes place during the learning of the second list seems evident. For a period of 24 hr. (under the conditions of the present experiment) the second-list responses are considerably stronger than the first-list response. By 48 hr. the two sets of responses appear to be of near equal strength. This was less clear in the present study than it was in the former study (**4**). This equalizing effect is empirically shown to be a function of the failure of the first list responses to decrease in frequency, coupled with the marked decrease in the frequency of the second list responses. It would be expected, of course, that the point at which the response strengths of the two lists become approximately equal will change as a function of certain manipulable variables, the degree of learning of the two lists suggesting itself most prominently.

It was noted that extraneous responses increased in over-all frequency with the passage of time. Is this increment to be interpreted as 'spontaneous recovery'? This is a plausible hypothesis if it be assumed that these response tendencies are the ones which would have been evoked by the stimuli prior to the learning of the two lists (excepting those cases in which stimulus and response generalization occurred). It is reasonable to expect that these adjectives used as stimuli had some responses, even though weak, already connected to them. Word association tests before the learning of the lists would have established whether or not these responses which came in with greater frequency with the passage of time were actually originally associated with the stimuli. If these associations were already formed, then during the learning of the lists these associations had

to be weakened (unlearned or extinguished) during the learning of the second list. These pre-formed associations may then recover with the passage of time. Indeed, the recovery of the pre-formed associations plus the recovery of the first list associations may constitute the basic mechanism which brings about interference with the second list associations and hence the decrease in frequency of those associations with the passage of time. In other words, these extra associations become a part of the process which is said normally to cause forgetting with the passage of time.

It should also be noted that the production of forgetting by the recovery of these previously learned associations would follow the paradigm of proactive inhibition. It is thus likely that the amount of forgetting caused by proactive inhibition is actually greater than the amount caused by retroactive inhibition. All previous associations formed during the life of an individual are potential interfering tendencies by proactive inhibition. Retroactive inhibition is limited only to the functioning of those responses acquired *after* the learning of the material to be recalled.

V. Summary

Twenty-four *S*s learned lists of 10 paired two-syllable adjectives by the anticipation method. For each of four experimental conditions, two lists having the A–B, A–C relationship were learned to a criterion of one perfect recitation. After the appropriate time interval *S* was presented with the common stimulus words and instructed to give the first of the two responses connected to a given stimulus which occurred to him. In four separate conditions this form of recall was given one min., five hr., 24 hr., and 48 hr. after learning the two lists. The results show:

1. As the time interval increased, a consistent decrease in the frequency of responses from the lists learned second.
2. No decrease in the frequency of the responses from the first list with the passage of time.
3. An increase in the frequency of responses from outside the lists with the passage of time.

These data are given as further support of the interpretation of unlearning of the first list as being similar to experimental extinction. The fact that no decrease in the effective strength of the first list responses takes place over 48 hr. suggests that a process running counter to the usual forgetting processes is present. It is suggested that this mechanism may be likened to spontaneous recovery.

(Manuscript received August 18, 1947)

References

1. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.
2. Melton, A. W., & von Lackum, W. J. Retroactive and proactive inhibition in retention; evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54**, 157–173.
3. Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**, No. 3.
4. Underwood, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38**, 29–38.

ANXIETY-PRODUCED INTERFERENCE IN SERIAL ROTE LEARNING WITH OBSERVATIONS ON ROTE LEARNING AFTER PARTIAL FRONTAL LOBECTOMY

BY ROBERT B. MALMO AND ABRAM AMSEL*

Introduction

With homogeneous items, serial position curves are characteristically bow-shaped with the peak toward right of center, giving a curve which is not symmetrical. This is the general finding of workers in the field of serial rote learning. With verbal lists, the bow-shaped curve has been obtained not only with single-item presentation using a memory drum, but it has also been obtained with the method of complete presentation, in which immediate recall was given in any sequence chosen by the subject. Moreover, this relationship has not been confined to verbal lists. It has also been observed in the learning of some mazes and in some rational learning problems (**15**).

A series of systematic investigations by Hovland (**6, 7, 8, 9, 10, 11**) have shown that quantitative variations in degree of bowing are determined by such factors as length of list, rate of presentation, and distribution of practice. Variations in the *shape* of the serial position curves have been demonstrated by McGourty (**16**) who found that serial position effects were functions of the construction of the lists and of method of measurement. Lists constructed in a way to distribute randomly the formal similarity among items in the list of consonant syllables yielded the usual bow-shaped serial position curve. When, however, the first three and last three items were formally similar, there appeared a curve which was not bow-shaped.

In experiments on intraserial inhibition, emphasis has, of course, necessarily been placed on the purely associative aspects of the learning. Investigation of factors related primarily to the learner rather than to such extrinsic factors as the material to be learned, or the conditions of learning has been a neglected subject in this field. Sears (**20**) has pointed out the particular need for experimental psychology to attack the problems of motivation, emotion, and conflict in relation to memory.

* Now at Department of Psychology, State University of Iowa. In 1946 A. Amsel submitted part of the material in this paper as a thesis entitled, "The Effect of Anxiety State upon Intraserial Interference in Rote Learning of Nonsense Syllables," to the Graduate School of McGill University in partial fulfilment of the M.A. degree.

In the present investigation, we undertook to study extra-task factors. Our purpose was to study the effect of 'clinical anxiety' upon learning. Our main lead came, of course, from the psychiatrists' observations that patients with 'anxiety states' often have 'poor memories.' From the experimental point of view it seemed heuristic to attempt to bring such a 'memory loss' within the framework of interference theory.

In so doing, we deemed it worth while to make the time-honored distinction between so-called 'organic' memory loss and 'functional' memory loss. In connection with the frontal lobe problem, one of the authors has presented the hypothesis that the 'organic' loss (from removal of the frontal association areas) amounts to increased susceptibility to associative interference (Malmo, **17**). In terms of intraserial inhibition, therefore, we predict that frontal lobe patients should present serial position curves which are significantly more bowed than those for unoperated subjects.

During the present investigation, whose main purpose was to study anxiety-produced interference, we were fortunately presented with the opportunity to study six bilateral frontal lobe cases. Rote learning data from these cases will be considered later, in relation to our prediction.

We may now proceed to a statement of the hypothesis, concerning anxiety-produced (or 'functional') memory loss. The main assumption is that the forgetfulness of the anxious patient is due to anxiety-produced interference (API) between the relevant response and some irrelevant response generated from the patient's continuous state of anxiety. Continuing along this line, it may reasonably be assumed that anxious preoccupation might yield to concentration upon the task at hand, or that degree of API could be diminished through the patient's own voluntary effort.

From clinical observations of patients with severe anxiety, further predictions concerning API in rote learning may be made. The most effective voluntary inhibition of the interfering anxiety-produced responses may be expected to occur following the presentation of the initial items, when failure or near-failure on these initial items should increase motivation. This reduction in API may be expected to last during the remainder of a single presentation. But during the rest interval between presentations, API may be expected to return in patients with severe anxiety states to nearly their original strengths.

These predictions are summarized in the following theorem which was tested by the present experiment. With homogeneous lists and factors related to intraserial inhibition constant from group to group, patients with 'clinical anxiety' will yield a serial position curve (for errors) which is significantly higher in its anterior portion, and not

significantly higher in its posterior portion than that of matched controls, without anxiety.

Experimental Procedure

The method of the experiment was to have *Ss* learn eight nonsense syllables by massed practice to a criterion of one errorless repetition. The syllables were learned in serial order by the anticipation method. The syllables were printed in 18-point block capitals on heavy paper. The individual syllables were then mounted on a continuous band. The syllables were exposed through a window one in. by .95 in. They were brightly illuminated without glare by a light behind the *S*. The syllables as presented could be seen by the *E* by means of a mirror arrangement. The E was completely cut off from the view of the *S* by a large screen, 36 in. high and 42 in. wide. The experiments were conducted in a quiet room. The syllables were presented at four-sec. intervals.[1] Between successive presentations of the list there was a six-sec. pause. Detailed records of all responses were kept on mimeographed forms.

The clinical group of *Ss* was composed of 16 non-psychotic patients in whom severe anxiety was noted as one of the most prominent symptoms. The diagnosis of anxiety was made by psychiatrists who had referred the patients to us for study.[2] The control group was composed of 16 *Ss* who matched the clinical group in age, sex, last grade completed in school, occupation and vocabulary. This group was drawn largely from the hospital staff.

In the clinical group age ranged from 22 to 45 years with a mean of 30 years. The range for the control group was from 20 to 43 years with a mean age of 29 years. The range for educational status was wide in both groups: from third grade to graduation from college in the clinical group, and from fourth grade to college graduation in the control group. Mean for last grade completed was approximately 10 for both groups. Revised Stanford-Binet Vocabulary Age levels ranged from Year XIV to Superior Adult III in both groups; and in both groups the average was at the level of Superior Adult II. About half the *Ss* in each group fell into classification III of the Minnesota Scale for Occupational Intelligence (3). About ⅜ fell into classes V and VI and only about ⅛ fell into the professional or semi-professional groups (Classes I and II).

On the first day all *Ss* learned two practice lists. Two days later, at the same time of day, each *S* learned a single list (experimental list 1). Two days following this, each *S* learned the final list (experimental list 2). The instructions, given at the start of the first session, were essentially the same as those used by Hull (14). Successive pairs of syllables had approximately the same sums of association frequency, as determined by Hull (13). The syllables in the experimental lists were systematically rotated so that each pair of syllables would appear in every possible place in the series when the subjects were taken as a group. Further, the units in list 2 were rotated on those in list 1 so that every pair of syllables in list 1 was presented in every possible combination with the pairs of syllables in list 2. Each control *S* learned exactly the same list (same syllables in same serial position) as did the clinical subject with whom he was paired.

Results

Effects of 'clinical anxiety' upon distribution of failures in various serial positions.—The results obtained in the present experiment are

[1] The syllable itself was exposed for two sec. only. Following this two-sec. exposure period a shutter blocked the syllable from view. The shutter remained down for two additional sec., during which time the *S* was permitted to anticipate the next syllable if he had not already done so. When the shutter rose again, the next syllable was in the window.

[2] It would be a gross over-simplification to regard the clinical group as distinctly homogeneous, because of the common factor of anxiety present in each case. Recently, Cameron (2) has presented systematic means for dividing the anxieties into different categories. It may be noted here that all but three of our 'clinical group' were assigned by a psychiatrist to the subgroup of 'facilitated anxiety state.' From this standpoint, our clinical group was relatively homogeneous. The small number of patients did not permit further breakdown, and for the purposes of the present investigation the presence of 'clinical anxiety' was considered sufficient justification for placing an *S* within the 'clinical group.'

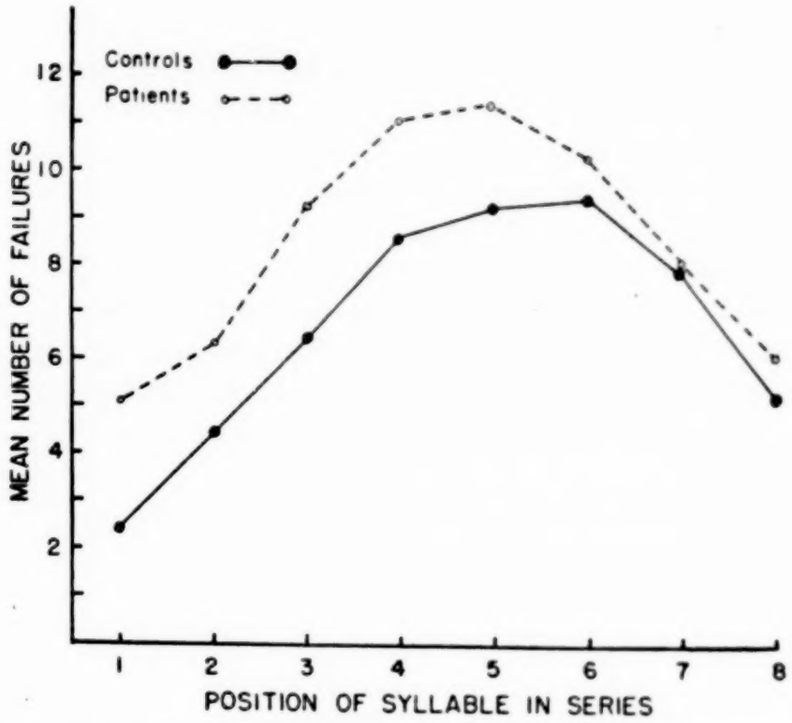

FIG. 1. Curves showing mean number of failures at various syllables during learning by patients and controls. Each point is based upon the mean of 32 learning scores.

in agreement with the predictions from our hypothesis concerning the operation of anxiety-produced interference in learning. In Fig. 1 a graph is presented, showing the mean number of failures for each syllable position in learning to the criterion of one errorless repetition

TABLE I

MEAN VALUES FOR THE DIFFERENT SYLLABLE POSITIONS IN LEARNING TO MASTERY BY CLINICAL AND CONTROL SUBJECTS. EACH MEAN BASED UPON 32 LEARNING SCORES

	Position of Syllable in Series							
	1	2	3	4	5	6	7	8
Mean number of failures								
Clinical group	5.1	6.3	9.2	11.1	11.4	10.3	8.1	6.1
Control group	2.4	4.4	6.4	8.6	9.2	9.4	7.9	5.2
Mean presentation at first success								
Clinical group	3.2	4.2	7.2	8.8	8.8	8.3	7.1	5.4
Control group	1.8	2.9	5.2	6.4	7.8	7.7	7.0	4.1
Mean presentation at last failure								
Clinical group	11.9	12.1	12.3	14.7	14.6	13.9	12.1	9.3
Control group	5.8	7.5	9.5	11.3	11.3	11.6	10.5	7.9
Mean range of oscillations								
Clinical group	8.9	8.0	5.1	6.1	5.9	5.9	5.2	4.4
Control group	4.5	4.8	4.7	5.2	3.3	4.2	3.8	4.1
Mean number of oscillations								
Clinical group	2.0	1.8	1.5	1.8	1.6	1.6	1.3	1.3
Control group	0.9	1.1	1.4	1.4	1.0	1.3	1.3	1.2

of the list. The first two rows of Table I present the values which were used in plotting these curves. The largest differences between the groups, measured in absolute terms, are manifested through the first five units of the serial list. The clinical group consistently made more errors on the first five units than did the control group. On the other hand, for the last three units of the list, performance was about the same from group to group.

An interesting comparison can be made between the present results and the results from a study by Hovland (11). In Hovland's experiment college students learned eight-unit lists of nonsense syllables under the two conditions of massed and distributed practice.

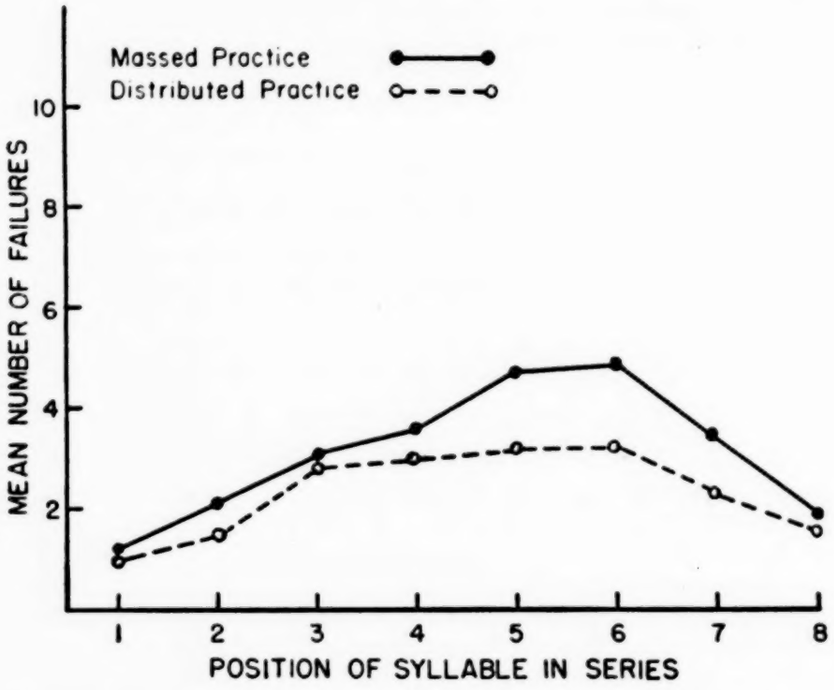

FIG. 2. Curves showing mean number of failures at various syllables during learning by massed and distributed practice. Each point is based upon the mean of 64 learning scores. Data from Hovland (11).

Conditions of learning were similar to those employed in the present experiment. Speed of learning was considerably higher for Hovland's *Ss* (see Fig. 2). The fact appears easily explainable in terms of their advantages of greater practice, higher intelligence and lower age.

Of particular note is the marked difference between the two pairs of curves with respect to syllable position number 6. Whereas in Hovland's experiment this was the place of widest divergence between the curves, in our experiment it represented one of the smallest differences. Conversely, the first serial position, which in our experiment yielded the largest difference, was the position which yielded the smallest difference in Hovland's experiment.

It appears from these comparisons that the method of serial learning affords a means of separating the effects of anxiety-produced interference and intraserial interference. The locus of the former appears to be confined primarily to the anterior portion of the series, whereas the latter occurs in the middle, or slightly to the right of center.

Table II presents the statistical reliabilities of the differences which were obtained. Only the anterior portion of the series yielded statistically significant differences, and the most reliable differences were obtained with the first syllable position. The posterior segment

TABLE II

RELIABILITY OF THE DIFFERENCES BETWEEN CLINICAL AND CONTROL GROUPS WITH RESPECT TO MEAN VALUES FOR ANTERIOR, MIDDLE, AND POSTERIOR SYLLABLE POSITIONS

	Position of Syllables in Series							
	First		First 2		Middle 4		Last 2	
	Mean Diff.	*P*	Mean Diff.	*P*	Mean Diff.	*P*	Mean Diff.	*P*
Mean number of failures	2.72	<.02	2.30	<.05	2.12	>.30	.55	>.60
Mean presentation at first success	1.34	<.02	1.31	<.05	1.54	>.20	.69	>.50
Mean presentation at last failure	6.03	<.02	5.17	<.05	2.93	>.20	1.53	>.40
Mean range of oscillations	4.34	<.05	3.81	<.05	1.40	>.20	.86	>.40
Mean number of oscillations	1.13	<.01	.89	<.05	.35	>.20	.00	—

yielded the smallest differences which were all very unreliable statistically. These statements hold true not only for the measure of mean number of failures, but for these other measures as well; mean presentation at first success, mean presentation at last failure, mean number and range of oscillations.

Effects of 'clinical anxiety' upon number and range of 'oscillations' at the threshold of recall.—An interesting difference between the clinical and control groups is revealed in the oscillations at the threshold of recall. Hull (**12**) has shown that *S*s will frequently respond successfully one or more times at a point in a rote series, and then fail at that particular syllable before another success occurs. Such a sequence of successes and failures he has called an 'oscillatory cycle' at the threshold of recall. From the first success to the last failure constitutes the entire range of oscillations. Hovland (**8**) and Shipley (**21**) have shown, independently, that the number of oscillations between the first success and the last failure resembles the serial position curves which depict only number of failures (or of correct responses). Hovland (**8**) has pointed out that the range of oscillations appears to be a good additional index of the amount of interference involved in the learning.

In Fig. 3 a graph is presented showing the mean range of oscillations, at each serial position, for the clinical and control groups. Also to be expected is the fact that the largest and most reliable differences between the groups occur at positions 1 and 2. For the first two positions, the difference between the groups is reliable at the five percent level. For these two positions, the difference in mean

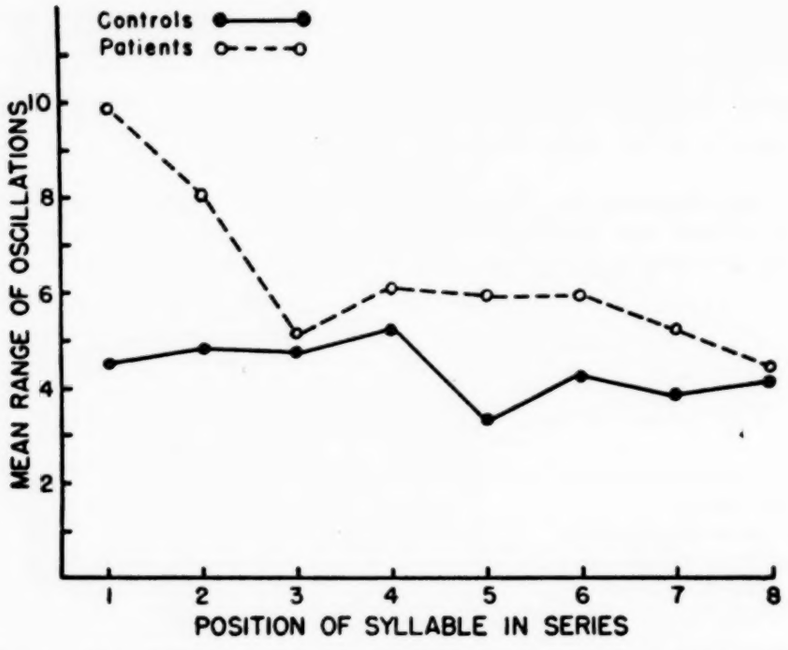

FIG. 3. Range of 'oscillations' at the threshold of recall. Curves showing mean number of presentations between first success and last failure for syllables in each position for learning by patients and controls. Each point is based upon the mean of 32 learning scores.

number of oscillations is also reliable at the five percent level (see Table II). Differences at the middle and posterior segments are lower, and they are not statistically reliable. The difference between the groups in terms of mean number of oscillations at the first serial position is reliable at the one percent level of confidence.[3]

Especially to be noted in our present results is the distinct tendency for the anxiety produced interference to become less as the *S* goes from the anterior to the posterior portion of the list. This tendency is particularly conspicuous through the first three serial positions. If the values for absolute number of oscillations are plotted (see Table I) the curve for the clinical group drops sharply

[3] Although our clinical group took slightly more trials to reach the criterion than did the control group, this difference was so small that we did not consider it necessary to apply an index of the sort used by Hull (12), in order to equate his clinical and control groups which differed markedly in time taken to learn lists of paired associates. Mean trials to learn the first practice lists were 21.37 and 18.68 for the clinical and control groups, respectively. Corresponding values for the second practice list, the first and second experimental lists were as follows: 16.38 and 13.75; 17.88 and 14.55; 17.38 and 13.81. In no case was the difference statistically significant: *P* was greater than .30 in every case.

from the first to the third position. In striking contrast to this, the curve for the control group is exactly opposite. It *rises* sharply, and the curves meet at position 3, forming a 'V.'

Range of oscillations for control *S*s also appears somewhat higher through the first half of the list. But through the first four positions, the curve lies parallel with the base-line, whereas the function for the clinical group quite clearly is a rapidly falling curve. The difference between groups for serial position number one is very significant statistically (see Table II).

The effect of continued practice in memorization upon anxiety-produced interference.—Anxiety-produced interference may reasonably be considered as a form of subject-produced distraction, to be distinguished from experimenter-produced distraction. The general finding from the more conventional distraction experiments has been that the interference value of the experimenter-produced distractions has decreased with the *S*'s continued practice on the central task. Clinical observations suggest that API may be a form of distraction which does not adapt so readily.

TABLE III

EFFECT OF PRACTICE UPON ANXIETY-PRODUCED INTERFERENCE AS MEASURED BY MEAN RATIO OF VALUE AT THE FIRST SYLLABLE POSITION TO THE AVERAGE OF THREE VALUES IN THE MIDDLE SYLLABLE POSITIONS*

Ratios of Value for Syllable Position Number 1 to Average of Values for Syllable Position Numbers 4, 5 and 6

	First List (Day 1)	Second List (Day 1)	Third List (Day 2)	Fourth List (Day 3)
Clinical group	.51	.56	.59	.48
Control group	.29	.37	.33	.14
Difference	.22	.19	.26	.34

* Because these ratios are influenced less by individual differences in speed of learning than are the absolute values, they seem more promising as indices, should the present method be developed as a test for use in clinical psychology. Another such index is available in the ratio of errors at position 1 to errors at the terminal positions (e.g. syllable position numbers 7 and 8).

List-to-list analysis of API provided a means of measuring adaptation in the present experiment. Table III presents the results of an analysis in which indices of API were employed. From the table there is the indication that while controls may have shown some adaptation, the patients did not. Comparing API indices for the first and fourth lists, the control value for the final list is less than one-half the value for the initial list; whereas the final value for the patient-group is nearly as high as their initial index.

Anxiety-produced end decrement

The phenomenon of 'end-spurt' in serial learning has been discussed by Melton (**19**) who showed that this phenomenon is an artifact in average memorization curves obtained by the Vincent-Kjerstad method. We observed no 'end-spurt' in the present study, but we did observe its opposite, namely, end-decrement which only occurred in the clinical group.

Fig. 4 presents the data graphically. Here, the number of trials required to reach successive criteria are plotted for the clinical and control groups. The curve for the control group shows negative

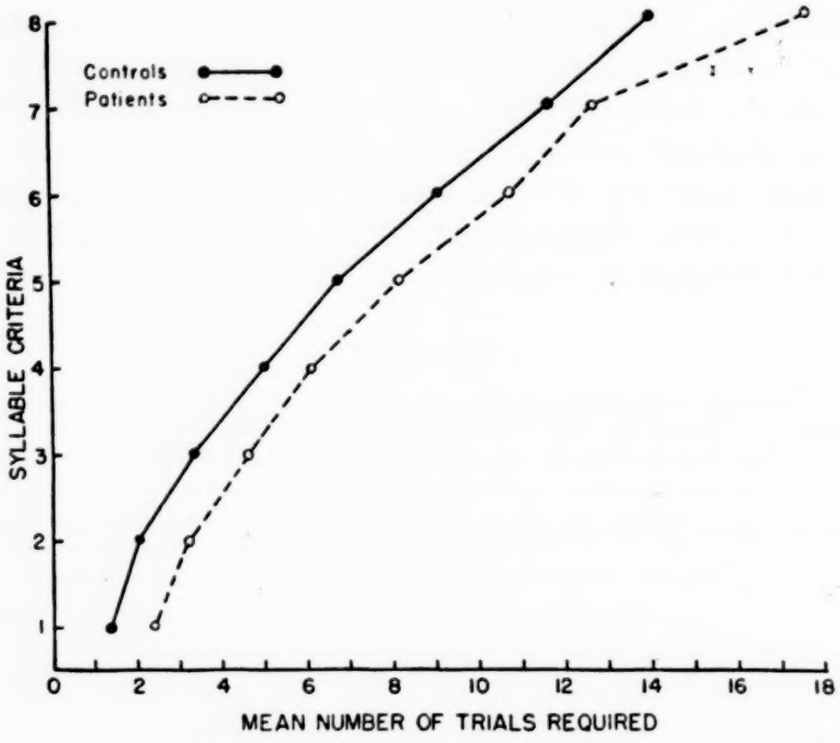

FIG. 4. Learning curves. Graphs showing mean number of trials to reach successive criteria of performance by patients and controls. Each point is based upon the mean of 32 learning scores.

acceleration with the latter part of the curve nearly linear. The curve for the clinical group is almost exactly parallel with that for the control group until the last one-eighth of the curve, where the curve for the clinical patients shows a terminal deceleration. The difference between the two groups in terms of number of additional trials required to advance from the criterion of seven correct to the criterion of eight correct is reliable at the five percent level of confidence. It should be emphasized that there is a small, *uniform*, insignificant difference between the groups in number of trials to go from each successive criterion to the next, up to the seven-syllable-correct criterion, and then a *large significant* difference between the groups in

terms of the number of trials to go from the seven- to the eight-syllable (or final) criterion. The absence of any tendency toward progressive divergence of the curves should be especially noted.

The fact that this end-decrement appeared in the curve for patients with 'clinical anxiety' without appearing in the control curve strongly suggests that the phenomenon was anxiety-produced. A possible explanation of this phenomenon has suggested itself to us following our own introspections during serial memorization. Late in the course of memorization there suddenly came to mind the realization that the task was almost completed. But we were by no means confident that we would get through the next trial without failure. The thought that the *E* might expect us to finish on the next trial occurred to us and there was the fear of not 'living up to' this expectation.[4] In the control group these fears may have been present, but if so, the interference from them was overcome within the time limit. In the clinical group, on the other hand, the fears may have been exaggerated to the point where a measurable end-decrement was produced.

Effects of frontal gyrectomy upon distribution of failures in various serial positions.—Because of the opportunities for analysis which it provides, serial learning appears a promising technique for distinguishing between interference which is primarily anxiety-produced and that which is 'organic' in the sense of involving damage to the brain. Our results make it appear that API is most pronounced in the initial segment of the series. Associative (intra-serial) interference, on the other hand, appears conspicuously in the middle segment. According to the hypothesis (Malmo, **17**) that frontal lobe patients are more susceptible to associative interference, they would be expected to show this greater susceptibility with more peaked serial position curves for error (that is, higher in the middle compared with curves for control subjects).

Six cases of bilateral frontal gyrectomy learned the same lists as did the other *S*s in this experiment.[5] The data from the gyrectomy group are presented in Fig. 5 along with the error curves for the 'anxiety' and control groups. The curve for one of the gyrectomy patients (H. M.) [6] was plotted separately because, on grounds other than performance in the present experiment, he appeared to be an

[4] A similar observation has been made in mirror drawing tests of anxiety where pre-final graphomotor disturbances were noted often in patients with anxiety conversion (**4**).

[5] The control group was chosen to match the 'anxiety' group, and it was not very well matched to the gyrectomy group. The control group had an advantage in terms of Vocabulary Score, age, and educational status. Excluding H. M., average Stanford Vocabulary Scores were 28 and 23; mean ages were 29 and 35; and mean grades were 10 and 4 (median grade for gyrectomy group was 7; one patient was taken out of school after her first year).

[6] A more detailed account of this case is to be published.

atypical case (18). The gyrectomy curve, compared with the curves for the 'non-organic' groups, has the appearance which would be predicted from the hypothesis that frontal lobe damage increases susceptibility to intra-serial interference. The greatest absolute differences occur in the central portion of the series.

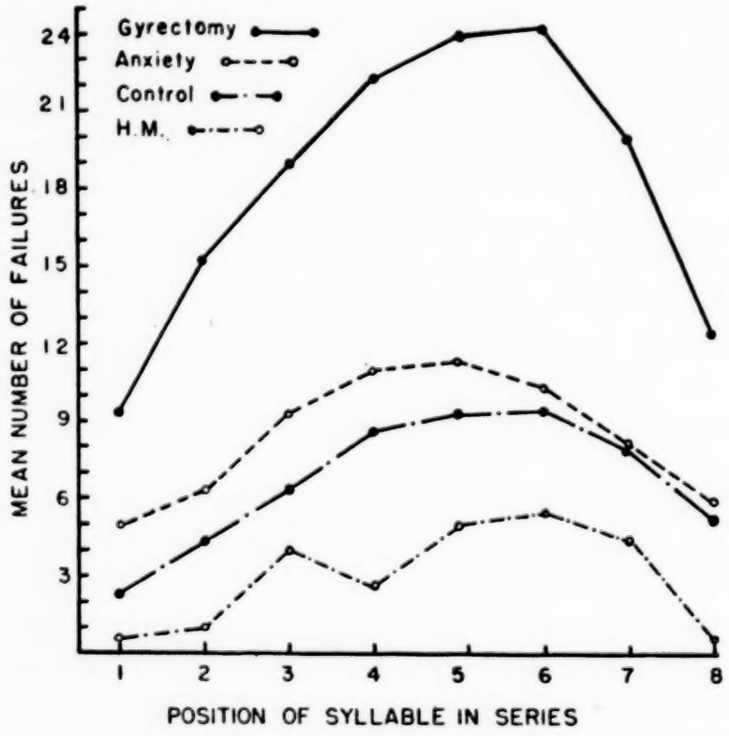

FIG. 5. Composite curves showing mean number of failures at various syllables during learning by 16 patients with 'clinical anxiety,' 16 controls, 5 gyrectomy patients, and H. M. (one gyrectomy patient who appeared atypical).

An important consideration is the fact that the gyrectomy group cannot be considered a non-anxious group. All of these patients had undergone operation for the relief of severe anxiety, and some residual anxiety definitely was present at the time they participated in this experiment. With this in mind, it should not be surprising that the gyrectomy group differed significantly from the control group with regard to failures in the anterior segment of the series (0.1 percent level). For the middle four positions, the difference was reliable at the two percent level of confidence. But, interestingly enough, the difference for failures in the last two syllable positions was not reliable at the five percent level. The absence of a reliable difference for the posterior segment suggests that the reliable differences may be a function more of anxiety and frontal lobe removal than of differences between the groups in terms of age, vocabulary and amount of schooling. Nevertheless, further research with matched groups is required before definite conclusions may be drawn concerning the

effects of frontal lobe removal upon degree of susceptibility to associative interference.

Discussion

Results from the present investigation were in line with predictions made regarding anxiety-produced interference in memorization. Interfering items of two sorts were distinguished: (1) intra-task (e.g. separate responses to two different nonsense syllables) and (2) extra-task (e.g. response to a nonsense syllable and an anxiety-produced response). The heterogeneous nature of the opposing items in API differentiates this particular kind of interference from that ordinarily observed and studied in the conventional rote-learning experiment. There the interfering items are supposed to be homogeneous in the sense of belonging to the intra-task category.

From the work on experimental rote learning it is known that degree of interference between intra-task items is a function of their similarity (**16**). It is important now to note that in the present case of API the degree of similarity between interfering items must be very low indeed. Some other dimension must be introduced in order to account for the effectiveness of interference in spite of the low degree of similarity. Intensity of response (related to severity of anxiety) is suggested.

There is the interesting possibility suggested here that the very dissimilarity between opposing items in API makes voluntary inhibition possible. This, of course, refers to the finding that clinical and control curves converged in the posterior segment. Assuming that the initial failures [7] were largely due to inhibition from interference we may propose an explanation for this 'recovery' in terms of voluntary inhibition of interfering responses. It seems reasonable to suppose that failure or near-failure on one of the early items might increase motivation toward the task. This increase in motivation might lead the *S* to inhibit the 'distracting' responses produced by his own anxiety.[8] Presumably such voluntary inhibition requires increased

[7] Our finding of increased frequency of error on the initial item in the series in anxiety appears related to the finding of Wechsler and Hartogs that initial graphomotor block is characteristic of patients suffering from anxiety neurosis (**22**).

[8] It is undoubtedly true that the rote learning situation is a stressful one, particularly for the anxiety neurotic. Some years ago one of us (R. B. M.) was engaged in taking rote learning data from college students. In the middle of his series of sessions, one subject sent us a note stating that he wished to cancel the remainder of his appointments. He complained that the sessions were making him so 'nervous' that he was not getting on with his girl. He concluded by saying that he "could not afford to have his nervous system shot to hell every day."

The rote learning test like the mirror-drawing test (**22**) is a situation which the *S* cannot meet with a highly canalized pattern of appropriate responses. From studies of neurotics in the mirror-drawing test, it is known that the stress from being required to deal with an unfamiliar

effort involving increased expenditure of energy, such as the increase which has been observed in distraction experiments. Although we did not have a measure of energy expenditure, it seems fairly safe to assume that something of the sort was involved here also. In accord with this view are the clinical findings that patients with severe anxiety generally show increased muscular tension. Tendon reflexes are increased, and tremors, especially of the hands, closed eyelids, or knees, are usually present. Cameron (1) has reasoned that fatigue which is usually present to a pathological degree in these patients, may be ascribable to increased energy expenditure in consequence of prolonged muscle tonus. There is the suggestion here that such increased tension reflects the patient's efforts to overcome anxiety-produced interference in everyday-life-situations.

From the experimental work on conditioning it is known that a conditioned response based on an involuntary unconditioned response, and itself relatively involuntary, may exhibit a certain amount of indirect susceptibility to verbal control (**5**, p. 266). Therefore it seems safe to assume that anxiety-produced responses, which are likewise relatively involuntary, may also be partially inhibited through voluntary effort, for brief intervals, under conditions of increased motivation.

Summary and Conclusions

1. Sixteen psychiatric patients, with severe anxiety as the predominant symptom, and sixteen control *Ss* learned lists of eight nonsense syllables by the anticipation method with massed practice at a four-second rate of syllable presentation. In addition, six cases of bilateral frontal gyrectomy followed the same procedure.

2. The results obtained in the present experiment were in agreement with the predictions from the following hypothesis: *The forgetfulness of the anxious patient is due to anxiety-produced interference between the relevant responses and the irrelevant responses generated out of the patient's anxiety-state. The interference value of the tendencies to irrelevant response may be decreased through voluntary effort on the part of the patient.*

3. From the results of the present study it appeared that the method of serial learning affords a means of separating the two effects of anxiety-produced interference and intra-serial interference. The

situation *accentuates* anxiety. It is probable that the rote learning test also accentuates anxiety. Level of anxiety is higher in anxious patients than in non-patients and it is generally higher in patients under stress than in patients not under stress. From the point of view of the effect of anxiety-produced interference upon learning, it is not crucial, however, to determine the degree to which the test situation has increased the basic anxiety of the patient.

locus of the former appeared confined, primarily, to the anterior portion of the series, whereas, as is well known, the maximum of the latter occurs in the middle, or slightly to the right of center.

4. Continued practice in memorization appeared not to reduce the degree of anxiety-produced interference in patients.

5. The fact that end-decrement appeared in the curve for patients with 'clinical anxiety' without appearing in the control curve strongly suggests that the phenomenon was anxiety-produced.

6. The gyrectomy curve, compared with the curves for the 'non-organic' groups, has the appearance which would be predicted from the hypothesis that frontal lobe damage increases susceptibility to intra-serial interference. The greatest absolute differences occurred in the central portion of the series. Further research with matched groups is required before definite conclusions may be drawn concerning the effects of frontal lobe removal upon degree of susceptibility to associative interference.

(Manuscript received August 22, 1947)

References

1. Cameron, D. E. Observations on the patterns of anxiety. *Amer. J. Psychiat.*, 1944, **101**, 36–41.
2. Cameron, D. E. Some relationships between excitement, depression and anxiety. *Amer. J. Psychiat.*, 1945, **102**, 385–394.
3. Goodenough, F. L., & Anderson, J. E. *Experimental child study.* New York: The Century Co., 1931.
4. Hartogs, R., & Wechsler, D. *Manual for the katoptographic measurement of anxiety.* (To be published)
5. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
6. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
7. Hovland, C. I. Experimental studies in rote-learning theory. II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **22**, 338–353.
8. Hovland, C. I. Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.
9. Hovland, C. I. Experimental studies in rote-learning theory. IV. Comparison of reminiscence in serial and paired-associate learning. *J. exp. Psychol.*, 1939, **24**, 466–484.
10. Hovland, C. I. Experimental studies in rote-learning theory. V. Comparison of distribution of practice in serial and paired-associate learning. *J. exp. Psychol.*, 1939, **25**, 622–633.
11. Hovland, C. I. Experimental studies in rote-learning theory. VII. Distribution of practice with varying lengths of list. *J. exp. Psychol.*, 1940, **27**, 271–284.
12. Hull, C. L. The formation and retention of associations among the insane. *Amer. J. Psychol.*, 1917, **28**, 419–435.
13. Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.
14. Hull, C. L. The influence of caffeine and other factors on certain phenomena of rote learning. *J. gen. Psychol.*, 1935, **13**, 249–274.
15. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942. (See esp. pp. 98 ff.)

16. McGourty, M. C. Serial position effects in learning as a function of interfering associations. Unpublished Master's thesis, Univ. Iowa, 1940.
17. Malmo, R. B. Interference factors in delayed response in monkeys after removal of frontal lobes. *J. Neurophysiol.*, 1942, **5**, 295–308.
18. Malmo, R. B. Personality factors in intellectual impairment following bilateral frontal gyrectomy. *Amer. Psychologist*, 1946, **1**, 267–268.
19. Melton, A. W. The end-spurt in memorization curves as an artifact of the averaging of individual curves. *Psychol. Monog.*, 1936, **47**, 119–134.
20. Sears, R. R. Functional abnormalities of memory with special reference to amnesia. *Psychol. Bull.*, 1936, **33**, 229–274.
21. Shipley, W. C. The effect of a short rest pause on retention in rote series of different lengths. *J. gen. Psychol.*, 1939, **21**, 99–117.
22. Wechsler, D., & Hartogs, R. The clinical measurement of anxiety. *Psychiat. Quart.*, 1945, **19**, 618–635.

STUDIES IN SPACE ORIENTATION. II. PERCEPTION OF THE UPRIGHT WITH DISPLACED VISUAL FIELDS AND WITH BODY TILTED

BY S. E. ASCH
Swarthmore College
AND
H. A. WITKIN
Brooklyn College[1]

In the first paper of this series (1) the role of the visual framework in the perception of the upright was investigated. By means of a tilted mirror the visual and postural systems, which usually coincide, were separated by a measurable amount. It was demonstrated that under these conditions the perception of the upright was determined in a fundamental way by the visual framework.

For the purpose of further investigation, the tilted mirror presents certain shortcomings. It is relatively difficult to vary the relation between the position of the observer and the mirror. Also, the mirror permits the observation of only forward and backward tilts of the field. For these reasons it was decided to abandon further observations with the mirror, and to work with a room that was actually tilted. The basic feature of the new procedure was for the observer to view a tilted room and to adjust a movable rod contained within it to the true vertical and horizontal.

EXPERIMENT I

Procedure

The tiled room, which was constructed of unpainted plywood, consisted of three walls and a floor, four feet wide by four feet deep by six feet high. It was thus open at the front and lacked a ceiling. Against the back wall of this room were placed a chair and a table. The latter had, in addition to a top, a lower shelf on which rested a book. Also, on the interior surface of the rear wall, directly above the pivot point of the rod, was a copy of the Boring 'wife-mother-in-law' reversible figure. Directly above these objects was mounted a horizontal rod, one in. wide and 39 in. long. The pivot point of the rod was located between the table and the chair. On the outside of the back wall of the room, attached to the rod, was a knob by means of which the rod could be rotated. A protractor system made it possible to determine the amount of rotation.

When the room had been arranged in this fashion, a horse was placed under its right edge, and it was tilted to the left at an angle of 22°.

In the front of this experimental room and at a distance of about six feet was placed a stand upon which a tube was mounted. The tube was circular in form, eight in. in diameter and 23 in. long, and was placed at such a height as to receive the *S*'s face, when he was standing erect.

[1] These experiments were performed in the Psychological Laboratories of Brooklyn College in 1942–43.

The field which the *S* was able to see as he looked through this tube consisted of the main contours of the tilted room, including the table, the chair, the picture and the rod. It included the lower part of the back wall, the rear portion of the floor, the back portions of the side walls, as well as the joints between the floor and the side and back walls. It is clear that in the perceived scene the vertical and horizontal were strongly represented. It was difficult to insure that all *S*s had precisely the same view, since the tube had to be adjusted to fit the height of the body. However, in each case it was made certain that the *S* could not see beyond the confines of the tiled room. If at any time an *S*'s view extended beyond the tilted room, the experiment was discontinued, and the data were not included.

The specific experimental procedures were as follows:

Situation a: The *S* was brought into the room blindfolded, and placed in position with his face in front of the tube. He was kept in this position, with his eyes closed, for three min. During this waiting period the object of the experiment was described to him as that of discovering the faces of the old woman and the young woman in the Boring figure. At the end of three min. the *S* was asked to open his eyes, and four additional min. were allowed in which to find the faces. If this task was accomplished in a shorter time, other tasks relating to the picture were given to fill in the four-min. period.

This period of preliminary adaptation was introduced to permit the normal visual framework to 'wear off,' and to provide an opportunity for the new field to exert its maximal influence. The reversible-figure task was introduced in order to diminish the tedium of the waiting period, and also to direct the *S* to the field with the help of an interesting task.

Following the four-min. period the *S* was informed that another kind of test was to be made, and the following instructions were given: "I am now going to move the rod in the room out of place, and I will rotate it slowly until you are satisfied that I have brought it into the horizontal position. When you feel certain that it is parallel to the ground on which you are standing and to the ceiling above you, tell me to stop." The *E* then moved the rod, by turning the knob on the outside of the rear wall of the room, and brought it to the near-vertical position. Then the rod was slowly brought back toward the horizontal until the *S*, satisfied that it was in the true horizontal position, asked that the movement be stopped. The position of the rod was taken on the protractor at this point. Following this first determination, the *S* was required to establish the vertical. In alternation two additional horizontals and two additional verticals were obtained, giving a total of three tests for each position. It will be noted that the instructions were to set the rod parallel to the walls of the room in which the *S* was standing (which, of course, he could not see).

Following these determinations, the *S* was questioned in detail concerning the basis of his judgments. Particular care was taken to ascertain that he had followed the directions for adjusting the rod. Finally, the tube was removed, and the *S* was permitted a brief examination of the tilted room from the same distance. He was questioned concerning any changes in the appearance of the room following the removal of the tube.

Situation b: The *S* was now brought to a position directly in front of the tilted room, so that his legs actually touched its sloping floor. A one-min. inspection period of the room from this position was allowed, following which the *S* was asked to establish the horizontal and vertical. The procedure employed in *Situation a* was followed here in every detail, three horizontals and three vertical determinations being obtained in alternate order. In this situation, it should be noted, the *S* had full knowledge about the objective tilt of the room, having had the opportunity to observe it against the background of the upright laboratory at the conclusion of testing in *Situation a.* Also he was able to see the entire interior of the room in the present situation. However, he could not see the surrounding laboratory, except in a very peripheral way, unless he deliberately raised his head. This, however, he was specifically instructed not to do. At the conclusion, the *S* was again questioned as to the basis of his judgments.

Situation c: The *S* was again placed at a distance of about six feet from in front of the tilted experimental room. The arrangement was identical with that of *Situation a,* except that no tube was employed, so that the *S* could view the tilted room, together with the surrounding upright room. The latter contained lines on the floor, pipes along the ceiling, the ledge of a blackboard, and many other prominent verticals and horizontals. Thus two visual frames of reference were available to the *S* in adjusting the rod—that of the tilted room and that of the upright laboratory. Again three horizontals and three verticals were obtained, according to the procedure employed in the two previous situations.

To summarize the salient features of the three situations:

In each variation the *S* had to establish the visual vertical and horizontal within a tilted visual field. In Situations *a* and *b* his view was restricted to the tilted framework, while in Situation *c* the surrounding upright field was also available. This upright field, however, was more remote from the rod than the tilted field. Further, in Situation *a* the *S* had no information about the true situation, having been brought in blindfolded. In both Situations *b* and *c*, however, the *S* had an opportunity to survey the entire scene, to gauge the amount of tilt of the small experimental room, and even to learn of the errors made in Situation *a*. Finally, whereas in Situation *a* a rather long period of adaptation was provided, to permit the fullest 'taking over' of the tilted framework, the adaptation period in Situations *b* and *c* was relatively short.

Seventy-six Brooklyn College students, both men and women, were tested in the three situations.

Results

The results show that in only five of the 456 judgments did the *S*s establish the H (horizontal) or V (vertical) accurately. These were all obtained in Situation *c*, when the *S* was able to see the outer upright laboratory. In each of the 451 remaining judgments an error was made, and in every instance the error was in the direction of the tilt of the room.

Mean scores for each of the three experimental situations are presented in Table I.

TABLE I

Mean Deviations in Degrees of the Obtained H and V from the True H and V for the Situations of Experiment I

Since each of the 76 *S*s tested made three H judgments and three V judgments, the values in the first two lines are based on an average of 228 scores, and the values in the last line on an average of 456 scores. All deviations were in the direction of the field so that larger scores reflect greater influence by the field.

	Situation *a*	Situation *b*	Situation *c*
Mean deviation of H judgments from true H	15.2°	15.7°	8.9°
Mean deviation of V judgments from true V	14.6°	14.9°	8.0°
Mean deviation of H + V judgments from true H and V	14.9°	15.3°	8.5°

Considering that the visual field was tilted by 22°, and the body by 0°, the judgments made in Situations *a* and *b* are closer to the axes of the tilted field than to those of the upright body. In Situation *a* the mean error for all judgments was 14.9°. For Situation *b* the mean error was 15.3°. In Situation *c*, where the *S* had available the many proper V and H lines of the outer laboratory, in addition to his own upright body, he was still significantly influenced by the more immediate tilted field, but to a lesser extent. Here the mean error was 8.5°.[2]

[2] The critical ratio of the difference between Situations *a* and *b*, using values based on H and V scores, is only 0.6. On the other hand, the differences between Situation *c* on the one hand and Situations *a* and *b* on the other are significant, the critical ratios being 9.1 and 9.9, respectively.

To permit a comparison of the individual performances, distribution curves have been constructed for each situation. These curves, which are presented in Fig. 1, are based on the combined H plus V scores of each *S*. It will be seen at once that in each of the three situations there is a wide range of deviations, reflecting marked indi-

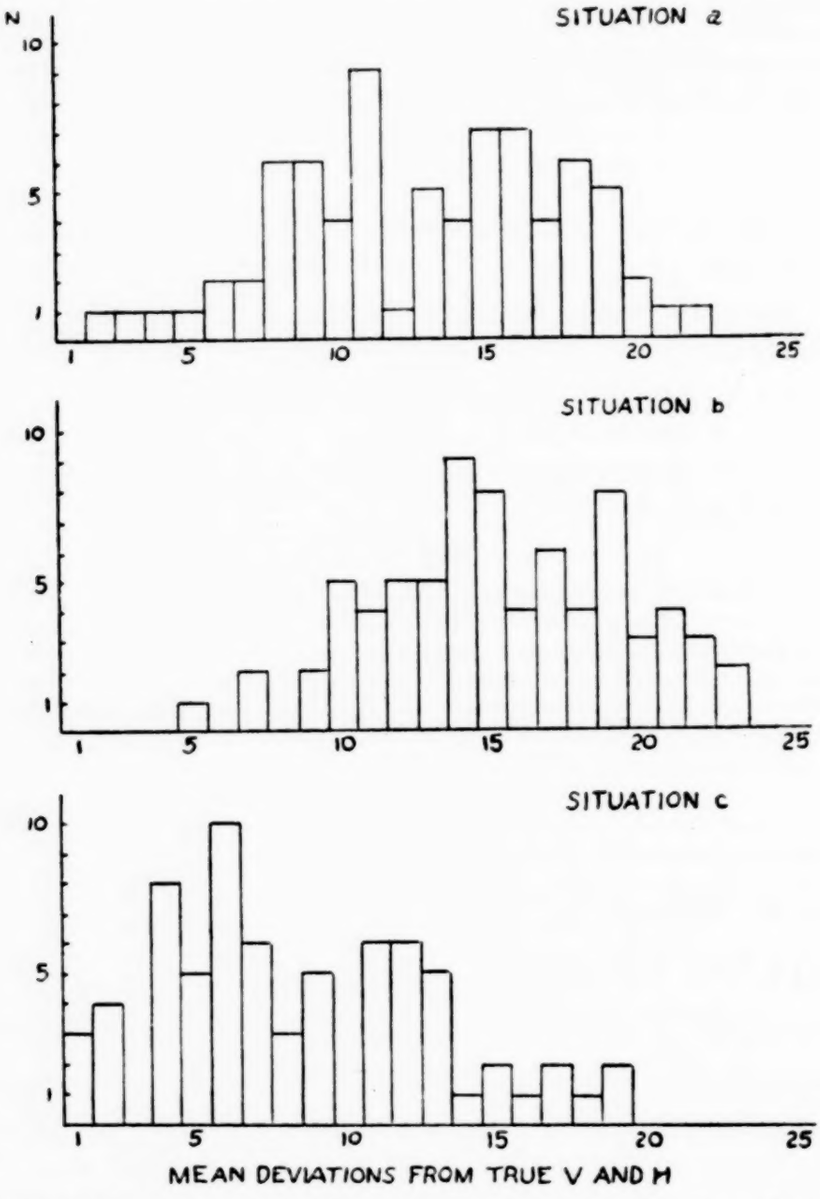

FIG. 1. Distributions of individual scores for each of the three situations of Experiment I. *N* is 76 in each situation

vidual differences. In Situation *a*, for example, there was an *S* at one extreme whose judgments deviated from the true H and V by an average of only 2°. At the other extreme was an *S* whose mean deviation reached 22°, coinciding with the tilt of the room. In each of the other situations similarly wide differences in performance were found.

The wide range of scores noted in each of the three situations reflects considerable individual differences in the extent of dependence upon the immediate visual framework. It remains to be decided whether a given degree of dependence upon the field is a consistent characteristic of the individual's orientation. To establish this, it is first necessary to determine the consistency of successive judgments made by each *S* under a given condition. Computing the range of the three H judgments and the three V judgments made by each *S* in Situation *a*, we find a mean value of 2.6° for the H determinations and 2.0° for the V determinations. These small values demonstrate that successive judgments made by an *S* are very similar to each other. Entirely in accordance with this result is the finding that the H and V scores in Situation *a*, which may be taken as representative, correlate to the extent of +.92.

Further evidence of consistency of performance is found when scores for the three situations are intercorrelated. Computed on the basis of H plus V scores, these correlations are found to be as follows: (1) Situation *a* vs. Situation *b*, +.80; (2) Situation *a* vs. Situation *c*, +.63; (3) Situation *b* vs. Situation *c*, +.63. These values indicate that a person who gave large deviations under one condition tended to do so in the other two situations as well. This suggests that a given degree of dependence on the visual field tends to characterize a person's orientation under various conditions.

These different lines of evidence—the marked consistency in successive judgments of the V and H in a given situation; the high correlations been V and H judgments made under a given condition; and the high intercorrelations among the different situations—suggest that the judgments possess a common basis. That this basis lies in the visual framework seems most probable, though more conclusive evidence for this interpretation may be found in experiments to be reported subsequently on the effects of elimination of the visual framework.

EXPERIMENT II

In each of the experiments hitherto reported the *S* stood erect while estimating the H and V. The upright position, it may be supposed, permits the most direct reliance upon postural factors: it is sufficient to align the rod with the body in order to establish the V and to place it at right angles to the body in establishing the H. What will be the effect of tilting the observer [3] upon his perception

[3] Tilting the body does not alter postural cues in a basic way, since bodily adjustments continue to be made to the downward-acting pull of gravity. To accomplish a fundamental alteration in postural factors, it would be necessary actually to change the direction of the force acting on the body. This was done in an experiment to be reported in a later paper.

of the visual V and H? The present experiment was designed to answer this question.

Procedure

The *a* Situation of Experiment I was modified as follows: Instead of standing upright, the *S* was seated in a chair which was itself tilted. This chair, which was slightly elevated and also had a back slant of 6°, was tilted 24° to the left. It was thus displaced to the same side as the experimental room, only 2° more. Attached to this chair was a tube of adjustable height, to receive the *S*'s face. The *S* was brought into the room blindfolded, and was assisted into the tilted chair. The tube was then placed in position over his face. In all other regards the procedure in Situation *a* of Experiment I was followed in detail. (To distinguish the two situations, the present one will be designated as *Situation a'*.)

Following Situation *a'*, the *S*s were tested in the *b* and *c* situations of the previous experiment. During these tests the *S* stood erect. Since these tests were identical in every respect with *b* and *c* tests of the previous experiments, they will also be designated as situations 'b' and 'c.'

Thirty-three Brooklyn College students, both men and women, served as *S*s in this experiment.

Results

The results of this experiment are presented in Table II. It will be noted, first of all, that the mean deviations from the true V and H for Situation *a'* are extremely high (19.3° for the H, 19.5° for the V, and 19.4° for the two sets of judgments combined). These values

TABLE II

Results for Experiment II

The values represent mean deviations, in degrees, of the reported H and V from the true H and V. The scores in the first two lines are based on 198 judgments, and those in the last line on 396 judgments.

	Situation *a'*	Situation *b*	Situation *c*
Mean deviation of H judgments from true H	19.3°	14.2°	10.2°
Mean deviation of V judgments from true V	19.5°	13.1°	9.7°
Mean deviation of H + V judgments from true H and V	19.4°	13.7°	10.0°

are quite close to the 22° displacement of the tilted room, indicating that on the average the rod was almost aligned with the axes of the room. Also, the average values are considerably higher than those obtained in the corresponding situation of Experiment I, Situation *a*, when the *S* stood erect.

A further striking result is that in Situation *a'*, 12 of the 33 *S*s perceived the room as perfectly upright throughout. Since these *S*s fully adopted the framework of the tilted field, they simply aligned the rod with its main axes in setting it to the V or H. This acceptance of the field was never found in Situation *a* of Experiment I, where the *S* stood upright. There, every one of a group of 76 *S*s perceived some tilt in the room. Since the sole difference between

the two situations is one of body position, the results indicate that when his body is tilted the observer increases his reliance on the visual frame of reference.

As may be expected from the virtual coincidence of the perceived V and H with the axes of the tilted room, the range of individual differences in Situation *a'* is narrowed, the scores now being concen-

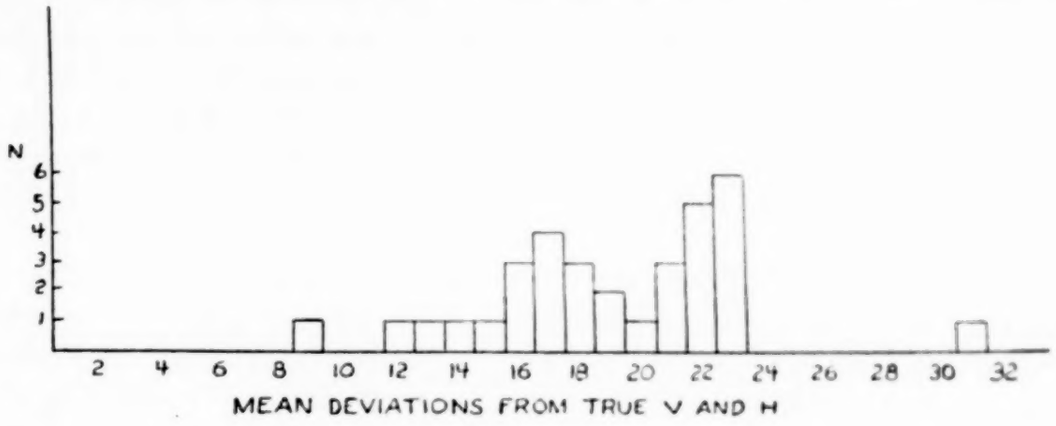

FIG. 2. Distribution of individual scores for situation *a'* of Experiment II. It will be noted that one *S* had a mean deviation of 31°. This *S* was one who had become very confused and displaced the rod beyond the tilted room rather than toward the upright.

trated at the upper end of the distribution. This becomes evident in an examination of Fig. 2, which contains the distribution of individual scores for this situation.

Situations *b* and *c* furnish results that approximate very closely those obtained with these situations in Experiment I.

EXPERIMENT III

PROCEDURE

In the variation just described, the *S* was tilted to the same side as the tilted field. In the present experiment he was tilted to the opposite side. Specifically, the room was displaced 22° to the right, and the *S*'s chair 24° to the left. Except for this difference in relation between body and room, the procedure was the same as in Situation *a* of Experiment I and Situation *a'* of Experiment II. Situations *b* and *c*, employed in each of the previous experiments, were omitted here. Seventeen *S*s were tested in this experiment.

RESULTS

For the group as a whole, the mean deviation of the H judgments from the true H was 20.0°, and the mean deviation of the V judgments from the true V 20.2°. In every instance the rod was displaced to the side of the tilt of the field. Thus, on the average, the judgments were within 2° of the axes of the tilted scene.

Of the 17 *S*s tested, 10 perceived the room as fully upright, and simply aligned the rod with its axes. Six other *S*s did recognize that

the room was tilted, but the perceived tilt was so small that it was necessary to make only slight allowances for it in adjusting the rod. Finally, the single remaining *S* became very confused, and reported the scene to be tilted in a direction *opposite* to its objective tilt.

It is of interest to compare the present findings with those of Situation *a'* of the preceding experiment, where the *S* was tilted *with* the field. The mean deviations are, in the present situation, somewhat larger. Also, the field was more often accepted as upright. Under the present conditions 59 percent (10 out of 17) of the *S*s perceived the room as fully upright, while only 36 percent (12 out of 33) did so in situation *a'* of Experiment II. There is evident an increased tendency to rely on the visual field when the body is tilted opposite to the field than when it is tilted with it.[4]

These results demonstrate even more strikingly than those of Experiment I that, within the limits employed, tilting of the body throws the *S* almost entirely onto the visual framework as a basis for judging the direction of a visual presented line, producing the virtually complete abandonment of postural factors as a basis for judgment.

Some Further Observations

The marked extent to which some *S*s were influenced by the visual field is dramatically confirmed in the comments they made during the period of observation, and in the course of subsequent questioning. Below are presented a few protocols of *S*s who departed extremely from the true V and H. These illustrate, in a way that the numerical results could not, the very real difficulty some *S*s encountered in making judgments under the conditions of these experiments.

The first *S* to be considered served in Experiment I. In Situations *a* and *b* she placed the V and H at 22°, simply aligning the rod with the axes of the tilted scene. At the conclusion of the test in Situation *a*, where her judgments were made quickly and with seeming confidence, she stated: "I didn't use the chair or table (in the tilted scene), because they seem to be at an incline. I made it straight up and down until it seemed in a straight line with me. That was for the V. For the H, I made a line with my own eyes." Despite an explicit effort to involve the body and to avoid

[4] It is likely that at least one basis for this difference may lie in the relative ease with which intellectual computations may be brought into play under each condition. When the body is felt to be tilted, and the scene is aligned with it, some people are able to deduce that, though the scene may appear perfectly upright, it cannot be possibly so. When body and room are tilted to opposite sides, however, an analysis of this kind is much more difficult. This is because such a difference between room and tilted body may arise when the room is tilted to the same side but less than the body, when the room is straight, and when the room is actually tilted to the opposite side. To choose among these alternatives, it is necessary to compute how much of the angle between body and room is due to the tilt of the body and how much, if any, to a possible tilt of the room. Thus, though in both instances the direct perceptual experience may be that the scene is upright, it is easier to deduce the true state of affairs when body and room are displaced to the same side than when they are displaced to opposite sides.

reference to the objects in the tilted field, this *S* was very much influenced by the field.

In Situation *c*, which included a full view of the outer upright room, this *S* on successive determinations adjusted the rod to within 2°, 5°, and 6° of the H of the tilted room, and to within 2°, 2°, and 3° of the V of the tilted room in establishing the H and V respectively. She reported that this situation was more difficult than Situations *a* and *b*, and actually required a considerably longer time to arrive at a decision. This situation should have been easier, as it was for most *S*s, for there was present an objectively upright frame of reference. She commented, on one of the trials: "It is much more difficult now, for I tried to compare it with the beams (of the ceiling) up there and with the walls (of the outer laboratory). It was very hard because the angle seemed so wrong. According to the (tilted) room the rod isn't vertical. For me it looks slightly to the left." It is clear from these statements that the *S* found this situation more confusing because of the 'contradiction' between two sources of evidence which she was unable to bring into relation. When the rod seemed straight in relation to the outer upright framework, it continued to appear very tilted in the tilted-scene context. (It is particularly striking that when the rod appeared so tilted that she had serious misgivings about it, it deviated only 2° from the axes of the room.) The paradoxically greater difficulty when there was more information available on which to base the judgment is thus due to the very compelling effect of the immediate visual field on her perception.

In short, when the tilted field alone was present, the *S* simply yielded to it, and her judgments were rapid and untroubled. When the upright field was introduced in the distance, she felt compelled to include it somehow. However, its influence was in the direction of taking the rod out of alignment with the scene. Yielding to this influence meant going counter to visual impressions, which for this *S* were extremely strong. The result was greater difficulty and disturbance in arriving at a judgment. What is at fault in this *S*'s preception is that, in consequence of her excessive dependence on the immediate visual field, she is unable to integrate all of the relevant factors in the situation. There is a strong adherence to a narrow frame of reference, where a proper judgment requires a broad frame of reference, encompassing the inner field, the outer field, and the body.

Difficulties of a similar nature were observed in another *S*, who served in Experiment I. When tested in Situation *a*, he essentially lined up the rod with the axes of the tilted scene on all trials but one. On this trial his judgment deviated by 6° from the axes of the scene. He gave his judgments quickly and confidently. At the end of testing in Situation *a*, he commented: "At first I noticed everything was on a slant. Therefore I had to let the rod go down a little more on the side toward which the slant was going. As I kept on looking the slant disappeared. [Specific experiments on the righting effect will be reported below.] It is a little slanted now, but before it was very slanted. In fact I even got a little dizzy looking at it." Most significant here are the righting of the scene with continued observation, indicating a tendency toward its full acceptance, and the dizziness

induced by viewing a tilted (and therefore, in a sense, unstable) though actually perfectly static scene.

In Situation *b*, this *S* made his judgments rapidly, again aligning the rod with the tilted scene on all trials. Despite this, he stated at the end of testing that he had used his body in making his determinations. He also noted that his body seemed to tilt with the scene. (In fact the *E* observed him to lean his head in the direction of tilt of the room and had to ask him to stand erect.) He stated further that he made use of the chair and the walls of the tilted scene. When asked if he had actually lined up the rod with the chair in the scene, he replied in the affirmative. It was pointed out to him that the chair was tilted, and he indicated that he was aware of it. When asked further if the rod was parallel to the walls of the outer building, he replied that it was. According to these statements, at each adjustment the rod was aligned with his upright body, with the upright building, and with the scene—which he knew to be tilted. It is entirely clear from other results that this kind of confusion was not due to a lack of understanding of the nature of the task. After some additional questioning along these lines, he came to the realization that his previous judgments must have been wrong, and having gained this insight, the test was repeated.

His V determinations now proved somewhat better, but he still deviated from the true V by 18°, 19°, and 19° on the three successive trials. In determining the H, however, he grew worse, and now tilted the rod 5° *beyond* the tile of the scene, or 27° from the true H. This he did on each of three successive trials. The basis of this extreme error becomes clear if one considers the dilemma the *S* faced. Under the influence of the *E*'s questioning, he had gained the insight that the rod must not be aligned with the tilted scene. There now arose the problem of determining in which direction it should be displaced. A decision on this point was more readily made for the V judgments, it being presumably easier to relate the rod to the upright body when the rod had to 'stand up' in about the same position as the body. For the H judgments, however, the decision was understandably a more difficult one. In the end an incorrect decision was made and the rod sent the wrong way. It is significant that in the retesting, when this type of error occurred, the *S* was typically more hesitant and required a longer time to make his judgments.

It is apparent that this *S* made an effort to avoid the use of the visual framework, but he had great difficulty in finding another basis for judgment. The body position was of some slight help in V judgments, but of no help in H judgments. Particularly in the case of the latter, the task became primarily one of getting the rod away from the scene. As far as appearances were concerned, it seems that the rod looked no more 'improper' when displaced to one side of the axes of the scene than when displaced to the other side. This is understandable in an *S* for whom the rod looked proper only when in line with the immediate scene. To appreciate how serious was the error of displacing the rod the wrong way, it is only necessary to recall that the *S* stood fully erect within four feet of the rod, and had full knowledge of the extent of tilt of the room.

The error of sending the rod the wrong way was observed in several other

Ss. The most striking instance occurred in an *S* who was receiving training directed at improving her performance. The training consisted of discussing with her the consequences of excessive dependence on the visual field, and of demonstrating to her the possibility of improvement by relating the judgments to the body position. In other words, this *S* was urged to adopt a 'body-conscious' attitude. During the original testing on Situation *b*, she did quite poorly, showing a strong tendency to go along with the visual field. After training, however, her performance became, on the whole, worse. Despite the instruction she went along with the tilted field on several trials even more than before training. In fact, on two trials she simply lined up the rod with the tilted scene, something she had not done before. At other times, when she did make an effort to 'resist' the scene, her judgments became quite wild, seemingly because she was unable to figure out which way the rod should be displaced to compensate for the tilt of the scene. Thus, on one trial where the rod was sent the wrong way, it was displaced 14° more than the tilted scene itself—that is, 36° from the true upright. Another result of the training, which really led to an effort at 'intellectualizing' about the situation, was to make for a highly variable performance. While on some trials there was a complete acceptance of the field, it was resisted on others; but as there was no stable basis for determining what form the resistance should take, the rod was brought nearer to the true upright at some times and removed very extremely from it at other times.

The effects of training were similar in Situation *c*, despite the presence of an upright field in addition to the upright body. On several trials the *S* again displaced the rod the wrong way. Thus, in one V determination, the rod was moved 20° beyond the tilted room itself, so that it deviated by 42° from the true upright. In this position the rod was accepted as parallel to her upright body. In another H determination, the rod was made to deviate by 32° from the upright. When after the latter adjustment the *S* was asked if the rod was now parallel with the floor she was standing on and the ceiling above, she replied in the affirmative. This she did after specifically *looking from each to the rod*. On other trials the *S* made errors in the opposed direction, over-compensating for the tilt of the scene and moving the rod by as much as 10° past the true H. On still other trials she came to within a few degrees of the true V and H. All told, the performance became highly variable and confused.

It is clear from the above protocols that in some persons the perception of the visual upright is fundamentally determined by the immediate visual framework. As a result there is great difficulty in including within the frame of reference non-visual experiences which may contradict the adopted frame of reference. The impression produced by the visual framework is so compelling that it is also strongly resistant to the influence of intellectual factors.

In contrast with the cases just described, the protocols of *Ss* whose quantitative results were closer to the objective V and H look rather different. The difference lies mainly in the absence of confusion and contradictions such as frequently characterized the quantitatively poorer

performances of other *S*s. A few excerpts from the protocols of one of the *S*s of Experiment I, whose numerical scores were particularly good, will illustrate this difference. In Situation *a*, the *S*'s H determinations deviated by an average of only 6° from the true H, and his V determinations by only 8° from the true V. He made his judgments fairly quickly, and in a precise, decisive manner. His comments at the conclusion of the test were as follows: "I was trying to visualize how I was standing. I held up my right hand. I couldn't see it, but it was an aid in getting the V. As for the H, I tried to disregard the rest of the environment and I tried to get a clear picture of when the rod would be in line with my eyes. I tried to figure out the angle of slope. I took the lower corner of the rod. I imagined how I was standing, and then I imagined how much I would have to bring the rod down to make it straight."

In Situation *b*, the mean deviation of his H judgments from the true H was 7°, and the mean deviation of his V judgments from the true V was 3°. His comments were as follows: "For the V, the first time it was just pure judgment. I went by the way I'm standing. I tried to keep myself erect. Then I tried to use the picture (in the tilted scene). When the right side of the rod touched the left extreme corner of the picture, I hesitated there, and considered the V. I noticed the angle it made with the picture the first time, and figured the angle should be the same later." (When asked whether the rod actually *looked* V in that position, or whether it was only that he had figured out that it must be V, he replied that it really looked V.) "For the H, the first time it was just a matter of judgment. I used my imagination. That time I noticed that one portion of the rod rested on the top hinge (in the corner of the tilted scene). Thereafter, I tried that to determine the H."

In Situation *c*, the *S*'s H judgments deviated from the true H by an average of 2°, and his V judgments by 1° from the true V. Now, before committing himself to a given judgment, he gave it lengthy consideration. His comments on how he arrived at his judgments were as follows: "I used the beam on the ceiling (of the laboratory). I visualized it, and then moved my hands along to see if it fitted. I also used the blackboard (on the wall of the laboratory). For the V I tried to imagine the rod parallel to the blackboard. As a check I tried to see if it was perpendicular to the ceiling and the floor. It was easier now than in the two previous situations."

In terms of his quantitative results, this *S* gave a considerably better performance than the ones considered above. His stated efforts to include a reference to his body were more successful than those of the other *S*s, who also attempted to judge the rod in relation to the body. There is no evidence whatever of the difficulties and confusion noted in the protocols cited previously. Further, this *S* seems to have taken account of wider aspects of the situation, such as the angle at which the rod should intersect various items in the room, the relation of the rod to his unseen hand, to his eye level, and so on. In general, this *S* displays a more discriminating and more 'workmanlike' approach to the situation.

The specific procedures employed by good *S*s and poor *S*s in establishing the upright have been investigated further. The fact that these procedures

can be determined only through the *S*'s own statements introduces certain limitations. First of all, the task of describing the basis of a perceptual experience is a new one for most. This often makes it difficult to describe clearly the method used in arriving at a judgment. Also, peripheral or even subliminal cues may be involved, and the inability of the *S* to report them results in an incomplete account of the method used. Despite these difficulties, an analysis of the *Ss*' reports seems worth while.

Such an analysis was carried out for Situation *a* of Experiment I, with the following results: (a) For 10 *Ss* whose sole method was to relate the rod to the visual field, by 'sending the rod the other way' from the tilt of the field to compensate for the tilt, the mean deviation of their judgments from the true V and H was 18.2°. (b) For 27 *Ss* whose sole method was to align the rod with the body, the mean deviation was 14.8°. (c) For eight *Ss* who made the rod cut the tube through which the scene was being viewed at angles corresponding to the V and H, the mean deviation was 13.5°. (d) Finally, for 23 *Ss* who used a combination of two or more of these methods, the mean deviation was 13.3°. Of the remaining eight *Ss*, two could not explain the basis of their judgments, one had the rod cut the picture in the tilted scene at a certain angle, and five related the rod to the imagined axes of the unseen outer laboratory. The last method was not very successful, the five *Ss* who used it having a mean deviation of 19.0°.

These data indicate that *Ss* who referred to their bodies alone did better than those who made their adjustments in relation to the visual field alone. The difference is in the direction expected on the assumption that a poor score results from excessive dependence on the visual field. It is significant that, although *Ss* who attempted to base judgments on the body did better, their rod adjustments were out of alignment with the body by an average of 14.8° when they perceived the rod and body as parallel.

It is also indicated that, among *Ss* who used a single method of judgment, those who had the rod cross the tube in a given fashion did better than those who attempted to relate the rod to the body or to the visual field. It is very likely that viewing the rod in relation to the tube leads to a better performance because the rod is thereby 'isolated' to an extent from its background. This undoubtedly reduces the influence of the tilted field to an even greater extent when the rod is aligned with the body; it is probable that when the rod is being related to the body it is still viewed as imbedded in the tilted scene.

Finally, the above data also show that *Ss* who employed several methods of judgment did better than those who employed one method alone. It is thus indicated again that *Ss* who take account of wider features of the situation do better than those whose view of it is more narrow.

EXPERIMENT IV

There is some evidence that with continuous inspection, a tilted field will appear less tilted for some *Ss*, and in some cases it will even right itself completely. Wertheimer (2) has reported that as the *S* continued to observe a tilted mirror scene, it gradually assumed a

normal position, and happenings within it that originally appeared tilted come to appear upright. It is important to establish definitely whether under conditions of prolonged observations the visual field becomes righted, or whether the postural factors assert themselves to a greater extent.

In the present experiment the *S*, from a tilted body position, observed the tilted room over an extended period of time. Measurements of the H and V were taken at regular intervals. A further feature of the situation was that the *S* had full knowledge of the tilt of the scene from the start of the observation.

Procedure

The *S*, with eyes open, was brought into the laboratory which contained the tilted room and a tilted chair. He was asked to survey the entire scene, his attention being directed particularly to the tilted room and the chair in which he was to be seated. After he had the opportunity to see the full tilt of the room, and the surroundings, he was seated in the tilted chair. As in Experiment III the chair was inclined 24° to the left, while the room he was viewing was tilted 22° to the right. There was therefore a disparity of 46° between the V of the body and the V of the room.

As before, the *S* was instructed to adjust the rod to the true V and H. The *first test* was given while the *S* was seated in the chair, with a full view of the tilted room and the surrounding laboratory. Immediately thereafter the tube was placed over the *S*'s face and another series of determinations was taken. This was the *second test*. The *S* was then questioned concerning the appearance of the various items in the tilted room. When three min. had elapsed, a *third test* was made. Finally, a *fourth test* was given after another interval of three min. Following the final test the *S* was questioned about his impressions of the scene. If his report did not include a statement concerning the apparent tilt of the scene, he was asked: "How does the tilt appear to you?"

At the end the *S* was informed that the tube would be removed and that he should observe carefully any changes following its removal. He was then allowed to examine the scene freely again.

Three V determinations and three H determinations were taken under each of the conditions described. Twenty-four *S*s, all college students, were tested in this situation.

Results

The mean deviations of the obtained V and H from the true V and H on each of the four tests are given in Table III, and the distributions of individual results for the four tests are presented in Fig. 3.

TABLE III

Mean Deviations in Degrees of the Obtained H and V from the True H and V with Prolonged Observation: Experiment IV

	Test 1		Test 2		Test 3		Test 4	
	V	H	V	H	V	H	V	H
Mean Deviations	10.5°	10.4°	15.1°	16.0°	17.6°	18.1°	17.4°	16.5°

The entries in Fig. 3 are based on the combined H plus V scores of each *S*. The following are the principal findings of the experiment:

1. With a full view of the tilted room and with full knowledge of the body position, the V and H judgments in Test 1 deviated markedly from the true V and H in the direction of the tilted visual field. The mean deviation was about half that of the tilt of the room, 10.4°

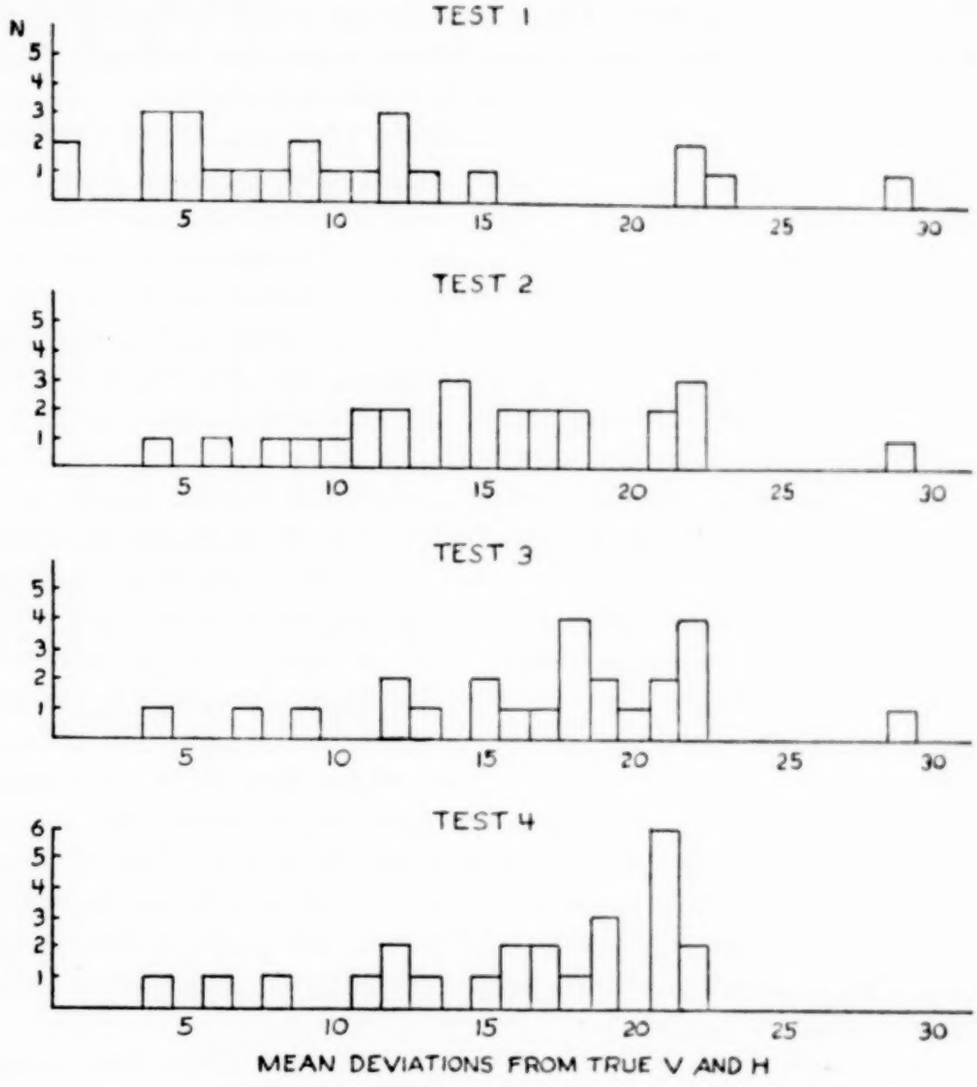

FIG. 3. Distribution of individual scores for each of the four tests of Experiment IV. *N* is 24

for the H and 10.5° for the V. At the same time the individual differences were striking, the deviations ranging from 1° to the full 22° tilt of the room. In one case, in fact, the rod was once displaced 15° further to the H than the room itself, or by 37° (see Fig. 3).

2. Immediately following the restriction of the *S*'s view to the tilted room along (Test 2), there occurred for the group as a whole an increase in the deviations toward the main axes of the tilted room.

The mean deviations rose to 16.0° for the H and to 15.1° for the V. This increase generally took place during the first minute of observation with the tube. For three *Ss* the field righted itself immediately, and their settings of the rod coincided either completely or nearly so with the tilted room. Five *Ss* in contrast showed no tendency toward an increase in their deviations; their performance was relatively uneven, and in some cases their judgments came closer to the true V and H. As may be seen in Fig. 3, for the group as a whole, the range of performances became more restricted, since the individual scores were displaced toward the upper end of the distribution.

3. Following three min. of observation with the tube (Test 3), the deviations increased still further, though the increase was smaller than from Test 1 to Test 2. The mean value of the deviations rose to 18.1° for the H and to 17.6° for the V. However, not all *Ss* contributed to this trend toward an increase. Those *Ss* who perceived a complete righting of the field in the preceding period did not of course add further to the rising deviations, nor did some *Ss* with a more fluctuating performance. Eleven *Ss* moved closer to the axes of the tilted room under this condition.

4. The changes observed during the fourth period (Test 4) were relatively slight. The mean deviation for H judgments dropped from the earlier value of 18.1° to 16.5°. (This drop is accounted for mainly by extreme fluctuations in the judgments of one *S*.) The V deviations dropped slightly, from 17.6° to 17.4°. Comparing individual results for Tests 3 and 4 we find that, for the H, the mean deviation increased in eight cases, remained the same in nine, and decreased in seven. For the V, the mean deviation increased in eight cases, remained the same in eight, and decreased in eight. Thus, as compared with the change from Test 2 to Test 3, there is found here a tendency for the reported values to remain unchanged. The absence of any significant difference between Tests 3 and 4 suggests that beyond a certain point, continued observation will not lead to any further righting of the visual scene.

At the conclusion of the final test, when the tube was removed from the *S*'s head, all *Ss* reported that the tilt had increased and also commented with surprise about the changed appearance of the tilted room. Most *Ss* stated that the tilted room now looked *smaller* and *nearer*. A typical comment was: "The whole perspective of the room seems changed. It is now much more tilted. It was further away before. The room seemed to have more depth before. Now it just seems a box." The same kind of report about the seeming change in the size and distance of the room upon removal of the tube was obtained in some of our other experimental situations.

We have already noted that *Ss* in this experiment differed mark-

edly in the extent to which their perception of the V and H was influenced by the surrounding tilted visual field. Fig. 3 illustrates this fact strikingly. Under each test condition the range of deviations obtained was very wide. Thus, on Test 1 the H deviations varied between 1° for one *S* and 37° for another. On Test 2, the range for the H was from 8° to 29°, on Test 3 from 7° to 29°, on Test 4 from 2° to 22°. Similarly the effect of restricting the visual field by means of the tube was markedly different for different *S*s. Some were greatly affected, while others were not affected at all. In the transition from the first to the second test, six *S*s showed increased deviations of 10° or more. At the other extreme there were six *S*s who showed no substantial change.

While the results indicate that prolonged observation leads generally to increased deviations, individual differences are again manifested. At the one extreme there were *S*s who 'deteriorated' markedly, in the sense that their deviations became larger and larger, while at the other extreme there were a few *S*s who were not consistent or who approached the true V and H more closely.

Perhaps the most striking result of the present experiment was the *complete righting* of the visual field for 11 of our *S*s. The reports of complete righting of the field, which the *S*s knew to be initially tilted, were obtained spontaneously rather than by direct questions. Following are some sample protocols: "It looks straight from where I am now. It feels straight . . . it feels as though I could walk into it. *I* would have to straighten up." One *S* conjectured that the room was tilted back to normal by the *E* while he was being distracted. "You distracted me with the moving rod. Now it (the room) seems to have been tilted back to such an extent that it is straight." Another *S* reported after three minutes: "I'd say, as I look at it now, it doesn't look as if there is any tilt at all. Absolutely straight, like any normal room. *I* feel slightly off balance." One *S* reported dizziness and nausea. "The room looks straight, but I feel very dizzy in it. I feel now that if I were to get off the chair and walk, I could walk right into the room. If I were to sit in this chair (the one in the tilted room) I'd feel perfectly normal." The reports concerning righting occurred within one to eight min. of the beginning of observations through the tube. (Since we waited for the *S* to report to us the occurrence of a righting effect, and since the report may not have been made until some time after the righting occurred, these time records are not a satisfactory index of the speed with which the righting of the field took place.)

It should be noted here that a few of the *S*s commented with reluctance on the righting effect, or showed embarrassment while reporting. Because they had initially seen the room as tilted, they

began to doubt the accuracy of their judgments. The great difficulty encountered by some of the *S*s in reaching a decision as to the proper location of the V and H is revealed in the following item from the protocols of one of our *S*s. After he had given judgments entirely in accordance with the main lines of the room, he made the statement: "Sometimes I think the room is upright, but sometimes it looks uncomfortable." He volunteered the information that in adjusting the rod he put it in line with the tilted room which seemed to him to be fairly upright. At this point it was suggested to him that he calculate the objective V. The result was that this *S* became uncertain of the very *direction* the true V should take, and finally decided on the wrong direction (i.e., he tilted the rod further than the room). He realized his difficulty, saying: "I really feel at a loss, I don't know which direction is right." The latter reaction is not, however, typical; most *S*s were quite sure of the correctness of the V and H which they reported. It should be noted further that as a rule the *S*s do not perceive the tilted field to move. They reported each time on the present tilt of the scene, comparing it to what they recalled of its previous tilt.

Discussion

The results of an earlier experiment (1) and those described here lead to the following conclusion concerning the way in which the visual upright is normally established: judgments of the visual upright are made by most persons in relation to a frame of reference provided by the main lines of the visual field. In this process postural experiences appear to have a secondary role in most people, and to function optimally only under limited conditions. Several lines of evidence support these generalizations:

1. When visual and postural factors are brought into conflict, as was done both here and in the tilted mirror experiment reported in a previous study, the perceived upright is typically nearer to the visual vertical than to the postural vertical.

2. Simply tilting the body increases the difficulty of referring to postural factors as a basis for judging the upright. Even more significant, as postural factors become less useful, the reliance upon the visual framework increases. In fact, under the conditions of the present investigation, most individuals went with the visual field either completely or almost completely when the body was tilted. This result is of particular interest when one considers that tilting the body does not alter at all the postural (gravitational) vertical, affecting only the ease with which its direction may be deduced. That so minor a change in the postural factors can so greatly impair their usefulness is evidence of their relatively limited role. These

factors achieve their maximum value only when the body is erect, and when no contradictory visual field is present.

3. Where tilting the body even slightly in the presence of a tilted visual field produces very serious errors in perception of the upright, the very same—or even more extreme—changes in body position in the presence of an upright field will not disturb perception at all. Thus, in an upright room a person may lean his head or body to one side,[5] or lie on his side, or even stand on his head, and still judge the direction of a visual line without any impairment of accuracy.

4. With prolonged observation, a tilted visual field has the tendency to right itself for most *S*s. Though both the visual and postural conditions remain constant during observation, the former becomes increasingly decisive for perception of the upright.

The evidence thus indicates that tilting the visual field, even with body erect, produces errors in the perception of the upright, in the direction of the tilt of the field. Conversely, tilting the body while the field remains upright does not disturb orientation at all, the upright being established with the usual precision. Tilting both the body and the field leads to the most serious errors, again in the direction of the tilt of the field. Finally, the tilted visual field tends to become phenomenally upright upon prolonged observation. These findings demonstrate how markedly perception of the visual upright is anchored to the visual field.

Not only do postural factors play a secondary role in perception of the upright, but it is even indicated that the maintenance of body position itself is affected by the surrounding visual field.[6] This was seen very strikingly in the tilted mirror situation and in the present tilted room situation where a number of *S*s were observed to fall to one side as they stepped back from the tilted scene. The loss of balance was apparently caused by the sudden shift from one visual frame of reference to another. In several instances the loss of balance was accompanied by dizziness. The statement made in one of these cases, "I'm all mixed up. I'm all dizzy," suggests that the sudden change of framework caused a momentary loss of bearings, which in turn induced the dizziness. The possibility of a causal relation between disorientation and illness is thus indicated. In this connection, it may also be reported that in Situation *b* of Experiments I and II there was manifested a strong tendency to lean the head or even the whole body in the direction of the tilt of the scene. While such effects were also noted in Situation *a*, they were more prominent

[5] In a control experiment it was found that every one of a small group of *S*s tested with body tilted and field upright was able to establish the V and H with complete accuracy.

[6] This phenomenon will be considered further in a later report, where experiments on the judgment of body position in a tilted visual field will be described.

in Situation *b*. The difference is due to the fact that in Situation *a* the *S* stood about six feet away from the tilted scene, while in Situation *b* he was right up against it. Being closer to the body, the tilted field began to exert a strong effect on body equilibrium itself.

Of great importance is the finding of large individual differences in the extent of dependence upon the visual field as a basis of orientation. These differences were noted both in the previous mirror experiment and in each of the situations described in the present report. Their significance is enhanced by the fact that there is considerable consistency in individual performance under different conditions. While the fact of a continuous distribution of scores speaks against the existence of sharply separated perceptual types, the individuals at the extremes do differ in a fundamental way. It seems highly significant that under the same field conditions a given line is perceived by one *S* as upright, by another as slightly tilted, and by still another as very tilted. The absence of an invariable relation between the structure of the field and the mode of perception clearly indicates that a full understanding of the perceptual experience requires consideration both of the outer field conditions and of conditions within the individual.

It has been established that the structure of the visual field exerts a marked effect on the perception of the upright. The evidence permits specific conclusions concerning the role of two major properties of the visual field.

1. The role of a unitary visual framework. The effect of the tilted visual field was markedly greater when it was the only field present. Its effectiveness decreased when the normal upright surroundings were also visible. Under the latter condition, judgment of the upright was influenced by both the inner and outer fields. As was demonstrated directly in Experiment IV (in the transition from Test 1 to Test 2), eliminating the outer upright field by restricting the view to the tilted field alone produced an immediate and strong shift in the perceived upright toward the vertical of the tilted field.

2. The role of articulation of the visual field. A more articulated visual field exerted a stronger effect than a relatively bare field. The evidence for this conclusion is derived from the tilted mirror experiment and from the present tilted room experiments. (a) The visual scene in the tilted mirror was found to be more effective when the *S* viewed it without a tube than when he viewed it with a tube. Without the tube, the *S* had a wider and more lifelike scene, and because he stood closer to it, he felt himself "surrounded" by the field, As a result he went increasingly with it, in judging the upright. (b) While some of the *S*s saw the mirror scene as fully upright from the outset, every *S* in the corresponding tilted-room situation (Situa-

tion *a* of Experiment I) saw the room as tilted. This occurred despite the greater tilt of the mirror. Since in both situations the *S* stood erect, this difference must be referred to the structural differences between the two visual fields. The mirror imaged a portion of a large room and of the scene outside the room, and provided a richly articulated field, with a very well defined framework. The tilted room, on the other hand, presented a field of much simpler structure. It consisted of some contours of the room, and the directionalized lines provided by only a few items—chair, table, and picture. The mirror scene clearly provided a 'stronger' visual framework, and this seems to account for its readier acceptance. This finding suggests that the specific structure of the visual field will determine how readily it will be adopted as a frame of reference for judging the upright.

Some of the results obtained indicate that the perceptual process investigated here is an autochthonous one which is not greatly affected by knowledge. Several lines of evidence support this conclusion: (1) In the tilted mirror situation removal of the tube caused the perceived vertical to shift further toward the tilted scene, though as a result of removing the tube, the *S* acquired knowledge about the tilt of the scene which was previously lacking. It might have been expected that knowledge of the given conditions would protect the *S* against error and aid him in correcting it. Not only was this not the case, but the results were quite in the opposite direction. The latter outcome resulted from a structural change in the situation, mainly a broadening of the visual field. This change in the situation determined the perception and not the knowledge gained. (2) Situations *a* and *b* of Experiment I produced essentially similar results while differing radically with regard to the knowledge available to the *S*. In Situation *a* the *S* was brought in blindfolded, and was given no knowledge of the tilt of the scene. In Situation *b* he had full advance knowledge of the tilt as well as of its magnitude. Again the factor of knowledge did not improve performance in Situation *b*. (3) Particularly impressive is the evidence concerning the righting of the visual field (Experiment IV). This effect occurred spontaneously despite the *S*'s initial knowledge about the tilt of the scene. For some, the righting seemed so real that they reached the conclusion that the *E* had actually straightened out the scene.

Further evidence of the limited role of knowledge comes from certain qualitative results obtained during these experiments.

First, in Situation *b* of Experiment I, the *E* at times moved the rod to the true upright after the *S* had adjusted it to what seemed to be the vertical. When informed that the rod was now objectively straight, many *S*s refused to believe it. This was the case even in Situation *c*, where the upright surroundings were visible. The rod

continued to appear as tilted as before. Indeed, the *E*s themselves had the same experience, despite many hundreds of observations and full knowledge of the true situation.

Secondly, in a number of cases the *S*s were asked, at the conclusion of the standard procedures, to step into the tilted room, and to stand upright in it while facing the back wall. It often happened that the room was then perceived as completely upright by the *S* despite his full knowledge of the tilt and despite the effort required to stand upright on the tilted floor. Even more striking were the *S*'s observations about his own position. Now the *S* reported that *his body appeared tilted.* The tilt, which was in a direction opposite to the objective tilt of the room, was most marked, and evoked deep astonishment The evidence is clear that the *S* tended to perceive his body in the same manner as he perceived the other objects in space, and that the position of his own body was in this case determined by the main lines of the visual field.

Thirdly, a somewhat different observation produced substantially similar results. Standing in front of the tilted room the *S* observed the *E* extending his arm about two feet into it. In his hands the *E* held one end of a plumb line. The *S* exclaimed with amazement that the freely suspended plumb line was tilted. The *E* also perceived the plumb as tilted despite many observations. If the plumb line was held fairly near to the side wall, some *S*s reported that it appeared as if it were deflected from the vertical position by a blast of air. It is seen from these results that knowledge can lead to a more critical attitude and produce a better approximation of reality, but it cannot greatly affect the influence of the field conditions. Indeed, in certain cases the effort to take into account the known facts impairs the performance; the errors may increase, the *S* may become confused, or he may infer certain additional conditions (e.g., that the *E* straightened the tilted room) that are contrary to fact. In the light of these observations, it seems hardly necessary to point out that the perception of the *S*s, though they involve extreme contradictions and though they fail to take into account extremely obvious features of the situation, cannot be regarded as unintelligent or unthinking.

Summary

To overcome certain disadvantages of the tilted-mirror situation as a technique for separating visual and postural determinants of the perceived upright, a new procedure was developed. The *S* was presented with a small tilted room, on the back wall of which was a rod which he had to adjust to the true upright. In a first test, judgments of the rod were obtained with body upright, and under three different

conditions of the field: (a) The *S*, standing at a distance from the tilted scene, viewed it through a tube which restricted his view to the interior of the scene. (b) The *S* stood directly in front of the scene without a tube. (c) The *S* stood at a distance from the scene, without a tube, so that he saw not only the tilted scene but the outer upright room as well. Under all three conditions the perceived vertical and horizontal were displaced significantly in the direction of the axes of the tilted scene. When an outer upright field was present, as in the third condition, the effect of the tilted scene upon the perceived upright diminished. In another experiment, judgments of the rod were obtained with body tilted, both to the same side as the field and to the opposite side. It was found that tilting the body resulted in an increased tendency to accept the tilted field as a basis for judging the upright. Not only were the perceived vertical and horizontal displaced further in the direction of the tilted scene, but a number of *S*s perceived the tilted scene as fully upright. In another experiment there was found a tendency for the tilted scene to right itself with prolonged observation. In some cases the righting was complete, so that at the end of the observation period, the tilted scene was perceived to be fully upright. Under all conditions, striking individual differences were found in the extent to which the perceived upright was affected by the surrounding tilted field. There is evidence of considerably consistency in a person's performance under the different conditions employed.

(Manuscript received for immediate publication April 26, 1948)

References

1. Asch, S. E., & Witkin, H. A. Studies in space orientation: I. Perception of the upright with displaced visual field. *J. exp. Psychol.*, 1948, **38**, 325–337.
2. Wertheimer, M. Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.*, 1912, **61**, 161–265.

PITCH CHARACTERISTICS OF SHORT TONES. II. PITCH AS A FUNCTION OF TONAL DURATION *

BY J. M. DOUGHTY AND W. R. GARNER

The Johns Hopkins University

INTRODUCTION

In the first experiment of this series (3) our interest lay in determining the shortest duration at which tones of various frequencies still had some pitch character to them. We determined two duration thresholds for pitch, representing the limits of pitch change from a 'click-pitch' to 'tone-pitch.' The 'click-pitch' threshold was the shortest duration at which a click was heard with pitch in it, and the 'tone-pitch' threshold was the point at which the sound became a tone and the pitch character changed little or not at all with further increases in duration. It was concluded that the 'click-pitch' threshold represented the shortest duration with which the ear can be stimulated in order for some pitch identification to be possible and the 'tone-pitch' threshold represented the shortest duration at which reasonably good pitch discrimination should be present.

Having established that pitch is present even in extremely short tones (on the order of four msec. for the middle range of frequencies and 11 msec. for low frequencies around 250 cps), the next step was to investigate whether or not the pitch of a tone changes as its duration is shortened. Certain it is, in view of previously published data on the problem of pitch discrimination as a function of tonal duration, that pitch *discrimination* becomes increasingly less accurate as the duration of a tone is shortened. Less certain, however, is the direction and the magnitude of the pitch change as a function of duration. The available literature presents contradictory reports on this relationship.

Bürck, Kotowski, and Lichte (2), for instance, state that the pitch of high tones is lowered and the pitch of low tones is raised by shortening the duration of the tone. Eckdahl and Stevens, as reported by Stevens and Davis (6, p. 101), on the other hand, conclude that the pitch of all tones (at least from 250 to 8000 cps) is lowered by decreasing their duration. The function for only one frequency, 1000 cps, is reported, however. According to this curve, a tone of 1000 cps lasting 10 msec. sounds equal in

* This research was carried out under Contract N5ori-166, Task Order I, between Special Devices Center, Office of Naval Research, and the Johns Hopkins University. This is Report 166-I-48 under that contract.

pitch to a tone of 842 cps lasting 1.5 sec. In other words, the apparent pitch of a 1000 cps tone is lowered considerably by shortening its duration to 10 msec. The data showing this relationship between duration and pitch were obtained from the observations of one *S*. The complete results of this experiment were never published.

Related to the problem with which out study is concerned are several experiments on pitch discrimination as a function of tonal duration. The data published by Békésy (1), based on observations from two *S*s, show that the differential limen for pitch increases as a function of shortening duration. The increase in the limen is gradual down to durations on the order of 100 msec., beyond which the curves rise at a rapidly increasing rate. For instance, in the case of an 800 cps tone, the difference-limen at the duration of 350 msec. is approximately .002, at 100 msec. it is about .004, and at 12 msec. it is about .098.

The problem of pitch discrimination as a function of tonal duration also concerned Turnbull (7). He presented his *S*s with a standard tone of a constant 500 msec. duration followed by another tone of a higher or lower frequency and of a known duration, ranging from 500 msec. to the shortest duration which allowed measurement of the DL. The *S* was required to judge whether the second tone was higher or lower in pitch than the standard tone. By having the *S*s make such judgments for a series of comparison tones of frequencies on either side of the standard, the *E* was able to determine the frequency increment or decrement at which the judgments of the *S*s showed 75 percent high or low respectively. The average of these two DL's, i.e., the least perceptible rise or drop in pitch, was the DL used by Turnbull in plotting his data. He justifies the use of the average DL on the grounds that he found no significant constant errors for pitch as a function of duration. His data agree with those reported in the Békésy study. There is a gradual increase in the DL from durations of 500 on down to 100 msec., beyond which the DL increases at an increasingly rapid rate.

In the cases we have reviewed above, studies aimed directly at the problem of pitch changes as a function of duration have been incomplete. Those dealing with the related problem of pitch discrimination as a function of tonal duration have given no information on the direction and magnitude of pitch changes, when such information should have been available in the data.

In the Turnbull study, it is quite possible that no consistent and reliable trends showed up under the method used, i.e., the method of constants. Our own experience in repeating Turnbull's work would tend to substantiate this to some degree, and a recent investigation by Koester and Schoenfeld (4) supports this possibility. They used a modified method of constants in their study of the effect of context upon judgments of pitch differences. Their results indicate that the degree and direction of pitch change is definitely affected by interaction effects operating in the method of constants. These interaction effects between standard and comparison stimuli might very well obscure the changes in pitch associated with the shortening of a tone's duration. About this more will be said in the discussion.

Purpose.—In the light of the foregoing discussion it is apparent that there was conflicting and incomplete evidence on the relation between tonal duration and pitch. These studies did point to the importance of duration as a variable, however, and our preliminary experiment on the duration thresholds for pitch suggested that reliable data for pitch judgments could be obtained even at extremely short durations.

In view of these considerations, the purpose of our experiment was twofold: (1) To resolve the contradictions found in the literature; and (2) to provide more complete information on the relation between tonal duration and pitch by using a wide range of frequencies and durations.

Experiment A: Method of Constants

Apparatus.—Our first attempt was essentially a repetition of the Turnbull experiment with some changes in apparatus and conditions. In our study, however, we were concerned with pitch changes rather than difference limens as a function of tonal duration. The essential difference in the apparatus was in the timing device. The timer in Experiment A was electronically controlled. The onset of both the standard and comparison tones, their duration, the interval between them and the repetition rate of the sequence of tones were all controlled by a specially designed electronic timing circuit. The timing device in the Turnbull study was mechanical. We feel that this difference did not influence the results. The variable frequencies of the comparison tone were manually controlled in both studies.

Conditions.—The conditions of our experiment differed in some respects from the Turnbull study. Standard frequencies of 250, 1000, and 4000 cps instead of 128, 1024, and 8192 were used. We employed two intensity levels, 90 db and 70 db *re* 0.0002 dyne/cm.^{2}, while Turnbull used a sensation level of 60 db throughout. In his study, the duration of the comparison tone took values from 500 msec. to the shortest duration which allowed measurement of the *S*'s DL; in our experiment the durations of the comparison tone were 6, 12, 25, 50, 100, and 200 msec. for 1000 cycles and 4000 cps and the same for 250 cps with the exception that 6 msec. was omitted.

All other aspects of the two experiments were essentially the same.

Our comparison frequencies for each standard were distributed symmetrically about the standard with one of them being equivalent to the standard. For the shorter durations, the comparison frequencies paired with a standard were spread over a wider range because of the larger difference-limen for pitch at these shorter durations. In most cases 11 comparison frequencies for each standard were selected. Where necessary, however, 13 were used.

Procedure.—In groups of three our *S*s were asked to listen, through earphones, to a series of tones consisting of a standard tone of 500 msec. duration, followed by an interval of silence of 250 msec., followed by the comparison tone of some predetermined duration and frequency. This sequence was repeated every four sec. The *S*s were instructed to write down on their recording sheets whether they thought the comparison tone was higher or lower in pitch than the standard tone, using the letters 'H' for high and 'L' for low. Six *S*s were used altogether.

During any experimental run only one standard was used, and with it each comparison tone was paired ten times. The order of presentation of pairs was randomized. Between each experimental run a short rest period was interpolated. No more than four runs were made during the experimental period of approximately one hour.

All observations were made in a sound-deadened room, and were made binaurally.

Treatment of data.—Using the method of least squares, an average psychometric function for any one set of conditions was determined by fitting a straight line to the combined data of all *S*s obtained under this set of conditions. The point of subjective equality was taken as the 50 percent point. In this way the pitch of 250, 1000 and 4000 cps tones at each of the selected durations was computed for the intensity levels of 90 and 70 db.

In plotting the data, the percentage of pitch loss or gain was used. These percentage values

were computed by substituting in the expression $\Delta f/F$, where Δf is the pitch loss or gain in cycles per second and F is the frequency of the standard tone.

An illustration from the results of this experiment will serve to clarify the reasons for stating that a particular frequency loses or gains a certain percentage of pitch at a particular duration. By computing the point of subjective equality as described above, it was found that a 1007.7 cps tone lasting for six msec. was judged to be equal in pitch to a 1000 cps tone sounding for 500 msec. (Fig. 1, upper 1000 cps curve). It was then assumed that a 1000 cps tone sounding for six msec. would be equivalent in pitch to a 992.3 cps tone lasting for 500 msec., or that it would lose 0.77 percent of its pitch by reducing its duration from 500 msec. to six msec. (Percentage pitch change = $\Delta f/F = -7.7$ cps/1000 cps $= -0.77$ percent.)

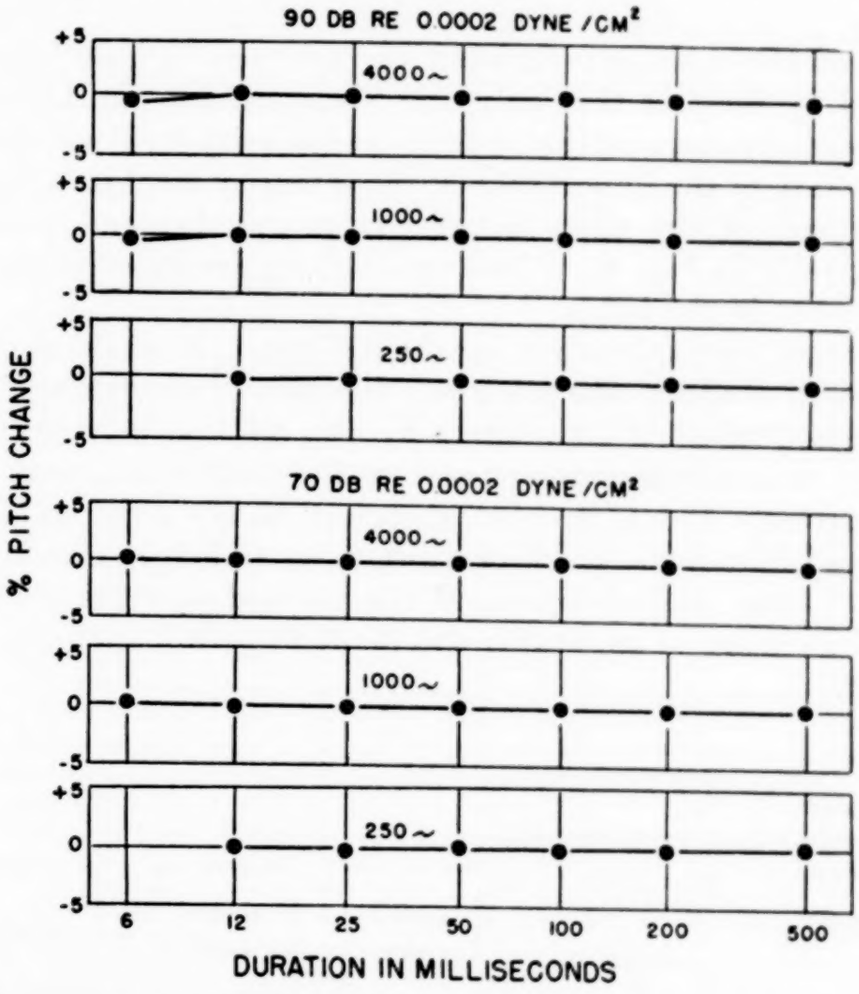

FIG. 1. The relation between tonal duration and changes in apparent pitch, as determined by the method of constants

The three upper curves are at the intensity level of 90 db *re* 0.0002 dyne/cm.² and the three lower curves at the intensity level of 70 db. Each curve is the average function for six *S*s.

Results: Method of Constants.—Fig. 1 shows the relation between tonal duration and percentage of pitch change for the intensity levels of 90 and 70 db. The abscissa for each curve shows the durations of the tone; the ordinate shows the percentage of pitch loss or gain for three frequencies of 250, 1000 and 4000. The upper three curves were obtained at an intensity level of 90 db and the lower three curves at an intensity level of 70 db.

Fig. 1 clearly indicates that there is a tendency, although only a slight tendency, for the three upper curves (90 db) to drop with shortened duration, indicating a loss of pitch. This drop is only noticeable at durations shorter than 12 msec. for the frequencies of 1000 and 4000 cps. The largest percentage of pitch loss was recorded for 1000 cps at six msec. This was only on the order of −0.8 percent, or approximately eight cps. This compares with a 15.8 percent drop in pitch recorded by Eckdahl and Stevens for a 1000 cps tone at 10 msec.

Part of the difference between the two sets of results might be explained by the fact that only one *S* was used in the Eckdahl and Stevens experiment, and that a different intensity level was employed. An indication of the importance of the intensity level is evident from a comparison of the three upper curves (90 db) and three lower curves (70 db) in Fig. 1. The drop at short durations in the former is not present in the latter.

Whatever the explanation for the difference, it would appear that if large pitch changes take place as a function of shortening duration, they are not revealed completely by the method of constants employed by us in this experiment and by Turnbull in his study of pitch discrimination. It was decided, therefore, to try a method in which the *S* could match the pitch of short tones to longer standard tones directly. Thus, in Experiment B the method of average error was used and a change in apparatus was made to permit such direct pitch matching.

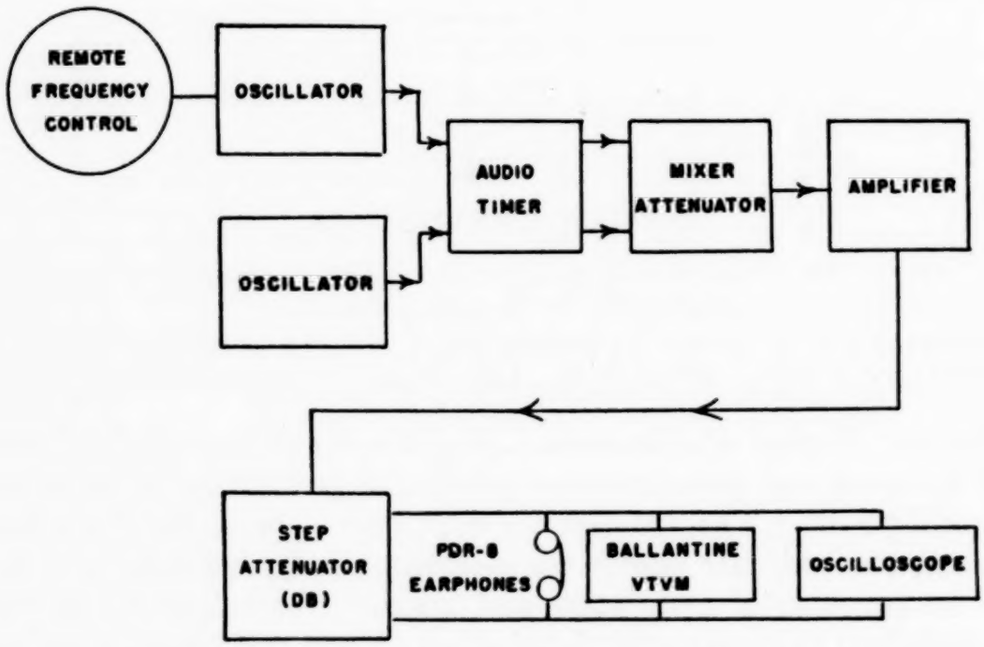

FIG. 2. Block diagram of the apparatus used in Experiment B

Experiment B: Method of Average Error

Apparatus.—Fig. 2 is a block diagram of the apparatus used in Experiment B. The change in apparatus mentioned consisted of installing a remote control which enabled the *S* to control the frequency of the comparison tone. In all other respects the apparatus was the same as that used in Experiment A.

Two oscillators were used: a Hewlett-Packard Audio Oscillator, Model 200A for the generation of the standard tone of 500 msec. duration, and a General Radio beat-frequency oscillator for the generation of the comparison tone.

The outputs of the two oscillators were led into the two channels of the audio timer which allowed for the control of the duration, separation, and repetition rate of the stimuli. From the timer the tones were conducted successively to a mixer attenuator, thence to an amplifier, a step attenuator, and finally into two sets of headphones: one set of Permoflux earphones (type PDR-8) for the *S*; and one set (type ANB-H-1A) for the *E*. These receivers were mounted in doughnut headsets, and binaural listening was used.

Connected in parallel with the earphones for purposes of calibration were a Ballantine Voltmeter and a Dumont (type 208-B) cathode-ray oscillograph.

The earphones which the *S* used had been calibrated by the Permoflux Company for operation into a six cc. coupler and they had a response which was reasonably flat between 100 and 8000 cps. For all calibrations of reference intensity the voltage across the phones was set at a level approximately equal to 110 db re 0.0002 dyne/cm.^{2}, as determined by the calibration curves.

Conditions.—The conditions for Experiment B were as follows: The standard tone of either 250, 1000, or 4000 cps was on for 500 msec., followed by a silent interval of 250 msec., after which the comparison tone of one of six durations—6, 12, 25, 50, 100, or 200 msec.—came on. The frequency of this comparison tone could be varied over a range of 0–20,000 cps by the *S* located in the sound-deadened room. This sequence of stimuli repeated itself every three sec.

Two intensity levels were used—90 and 70 db re 0.0002 dyne/cm.^{2} A small amount of white noise was also introduced to the listener's earphone to prevent the possibility that the *S* could detect weak beats from the slight electrical leakage of the audio timer. This noise was 50 db below the level necessary to mask the steady pure tone.

Procedure.—The same six *Ss* used in Experiment A were used in Experiment B. They were all male students between the ages of 22 and 28 years. In most of the experimental sessions at least two *Ss* were present. They alternated in making their judgments. They sat in a sound-deadened room, placed the earphones over their ears themselves so that they fitted comfortably, and at a signal from the *E*, who was in another room, proceeded to manipulate the remote control, which allowed them to alter the frequency of the comparison tone over the entire audible range of frequencies, until they had matched the standard and comparison tones in pitch.

Result: Method of Average Error. Effect of duration.—Figs. 3 and 4 show the relation between the duration of a tone and the change in apparent pitch as a function of this duration. So that the results of our two experiments might be more easily compared, Figs. 3 and 4 have been drawn to the same scale and with the same arrangements of variables on the axes as was Fig. 1. The abscissae show the duration of the tone in msec. and the ordinates show the percent of pitch change as determined from the expression $\Delta f/F$. Each point on the graphs in Figs. 3 and 4 represents the average of from 25–30 judgments. (No judgments were obtained at the duration of 500 msec. but it was assumed that no change in pitch occurs at this duration.) From two to five judgments were obtained from each *S* under each condition.

The shaded area around each of the curves represents the variability of the *Ss*' judgments as measured by the sigma of the individual judgments above the average.

Fig. 3 represents the changes taking place in the apparent pitch of a tone as the duration is varied with an intensity level of 90 db. It is apparent that considerable pitch change takes place toward the shorter durations. There is a decided tendency for all frequencies, i.e., 250, 1000, and 4000 cps to lose pitch as their durations are reduced to 12 and six msec. The greatest change, on the order of −4.0 percent, takes place at a frequency of 250 cps and a duration of six

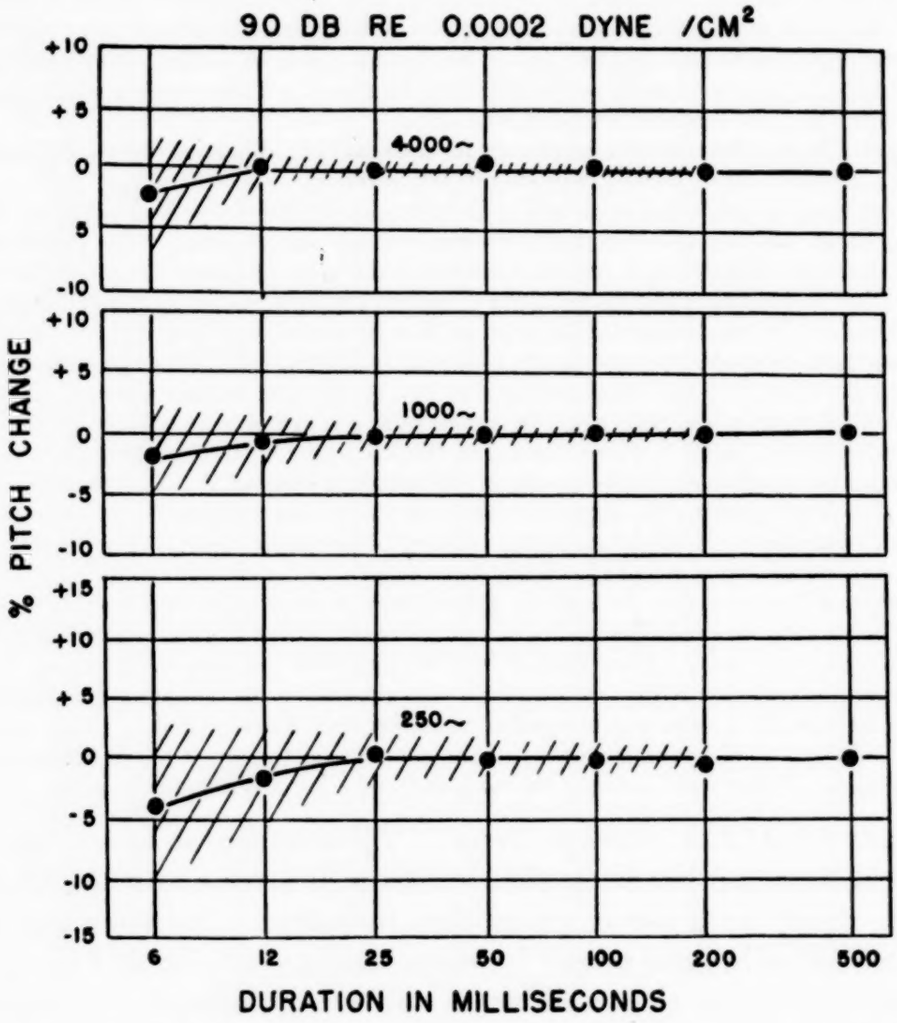

FIG. 3. The relation between tonal duration and changes in apparent pitch, as determined by the method of average error

The intensity level was 90 db *re* 0.0002 dyne/cm.² Each plotted point represents the average of 25–30 observations obtained from six *S*s. The shaded areas represent the sigma of the *S*'s observations about the average.

msec. About the same amount of change and in the same direction is common to both 1000 and 4000 cps. At these frequencies, the percent change is on the order of −2.0 percent at six msec.

These percentages are considerably larger than those reported for Experiment A. The reader will recall that the percentages of loss in that experiment were less than one percent at any frequency.

However, it is to be noted that both sets of data confirm each other in the direction of change. Fig. 1 and Figs. 3 and 4 agree in showing the trend toward a loss in pitch at the shorter durations, at least, in the case of 1000 and 4000 cps.

It is notable that the pitch of a tone remains relatively constant down to durations of 25 msec., being affected seriously only at durations considerably shorter.

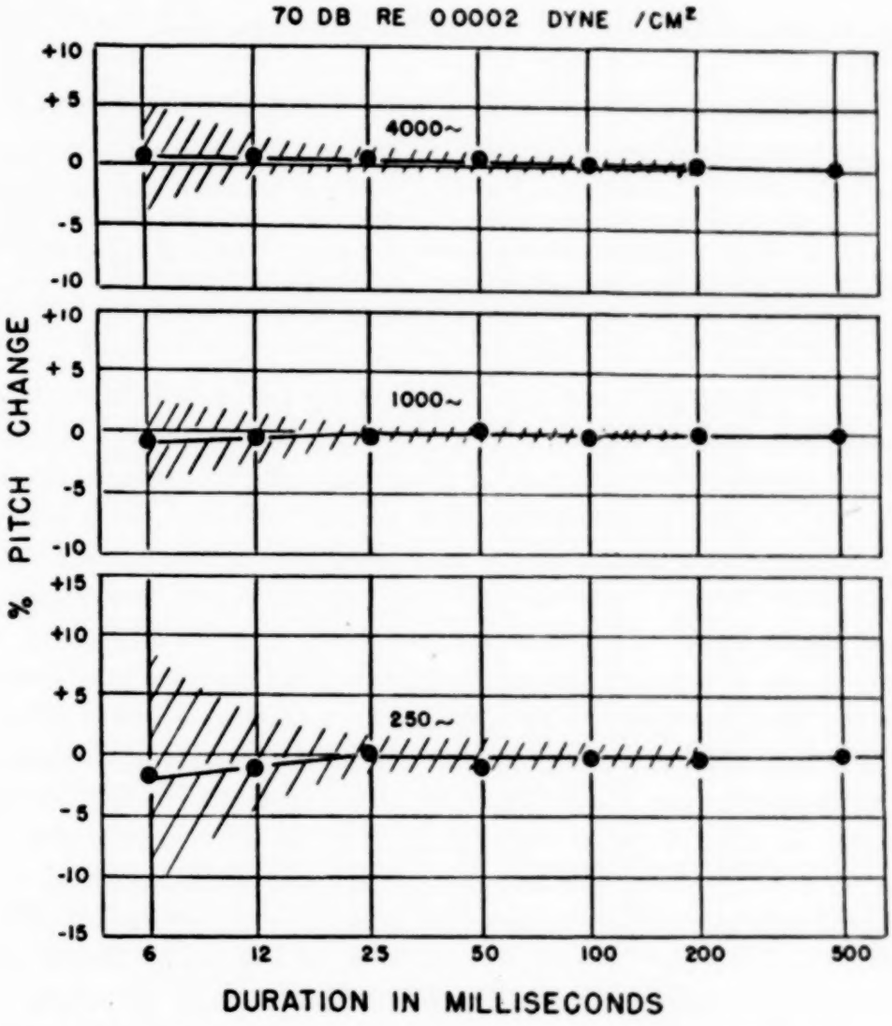

FIG. 4. The relation between tonal duration and changes in apparent pitch, as determined by the method of average error

The intensity level was 70 db *re* 0.0002 dyne/cm.² Each plotted point represents the average of 25–30 observations obtained from six *S*s. The shaded areas represent the sigma of the *S*'s observations about the average.

Effect of intensity.—Fig. 4 shows the results obtained when an intensity of 70 db was used. The striking fact about these curves is that the amount of pitch loss occurring at the extremely short durations is less than that shown at the higher intensity level of 90 db (Fig. 2). This is true for 250 and 1000 cps, and for 4000 cps there is a decided trend toward a gain in pitch as a function of shortening duration.

This same influence of intensity is evident when the three upper and three lower curves of Fig. 1 (Experiment A) are compared.

From these results it is obvious that both frequency and intensity interact in determining the direction and amount of pitch change as a function of tonal duration. For instance, the percentage of pitch loss for a 1000 cps tone at a duration of six msec. at an intensity level of 90 db is 1.8 percent; at 70 db the same tone has a pitch loss of only 0.8 percent. A 4000 cps tone of six msec. duration shows a 2.2 percent loss at 90 db and an 0.8 percent gain at 70 db.

Experiment C: Intensity and Frequency Effects

Apparatus and procedure.—The apparatus and procedure for Experiment C were the same as for Experiment B with the exception that only one *S* was used (not one of the original six).

Conditions.—Five different intensity levels were selected: 110, 90, 70, 50, and 30 db *re* 0.0002 dyne/cm.^{2}, and six different frequencies were used: 250, 500, 1000, 2000, 4000, and 8000 cps. The duration of the comparison tone was kept constant at 12 msec., and the standard tone was again 500 msec. Ten judgments per condition were obtained from this one *S*, to check further the interactions between intensity and frequency.

Results.—Fig. 5 shows what happens to the pitch of a tone of 12 msec. duration and of a given frequency as the intensity of that tone is varied. As the intensity is decreased, pitch change tends, successively, to become less negative (or more positive), to level off, and then to become more negative (or less positive) at very low intensities. This generalization is true for all frequencies with the exception of 4000 and 8000 cps. At these frequencies, the levelling off occurs, but there is no tendency toward negative values at low intensities. Noteworthy about 8000 cps, too, is the fact that at all intensity levels, even at the highest, the pitch change is in the positive direction.

For this particular observer, the pitch change for a 250 cps tone at 12 msec. is contrary to that indicated in our combined data in Figs. 2 and 3. That is, the average of six *S*s' judgments at 12 msec. for a tone of 250 cps shows a loss of pitch at both 90 and 70 db. The average of this one *S*s' judgments, shown in Fig. 5, at these intensity levels and at the same tonal duration show a gain in pitch. At all other frequencies, however, the judgments of this one *S* agree quite well with the average judgments of the six *S*s. We feel, then, that these data are in no real sense atypical. The previous observation about the effect of intensity on the pitch of short tones is confirmed, and the relation somewhat extended by observations from this one *S*.

Discussion

Comparison between Experimental Methods

A comparison between the results obtained with the method of constants (Experiment A) and the method of average error (Experi-

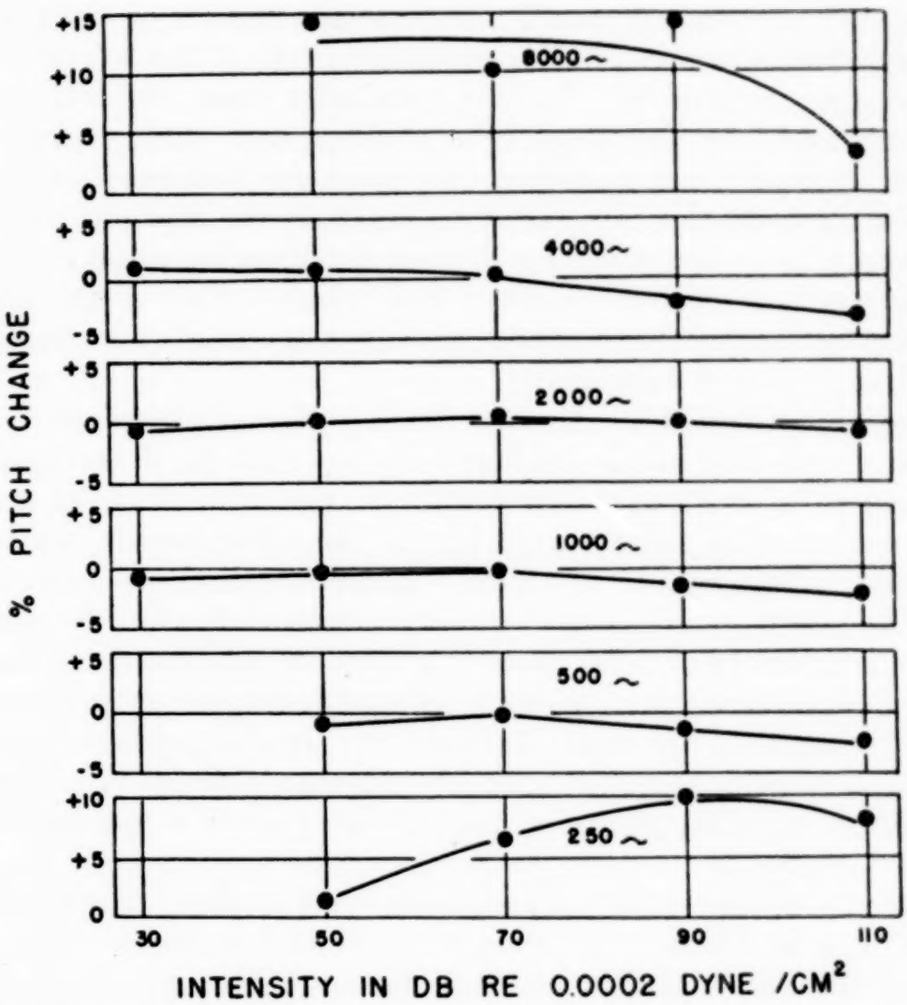

FIG. 5. The relation between intensity and changes in apparent pitch of 12 msec. tones

Each plotted point represents the average of 10 observations obtained from one *S*. The duration of the standard tone was 500 msec.

ment B) shows a difference in the functions representing the relation between tonal duration and pitch. This difference is chiefly in terms of the magnitude of the observed changes. Thus, from Fig. 1 (method of constants) it can be seen that the changes in pitch are never in excess of 1.0 percent at 90 db and that no changes in pitch appear to be present at 70 db. On the other hand, from Fig. 3 (method of average error) it can be seen that changes of between 2.0 and 4.0 percent occur at 90 db and from Fig. 4 (method of average error) that changes of between 1.0 and 2.0 percent are present at 70 db.

That different results can be obtained by using different psychophysical methods in investigating a given problem is nothing new. In a recent study Postman (5), for instance, shows that time-error functions for tonal stimuli obtained by the constants method are quite different from those obtained by a method of average error. The

problem for the experimentalist is to select the set of data, or that method, which gives the better representation of the relationship he is seeking to determine. In this particular case, we feel that the results obtained by the method of average error best represent the relation between tonal duration and pitch for two reasons: (1) The kind of judgment the *S* is asked to make in the method of average error is more relevant to the problem of determining pitch changes, or judgments of subjective equality; and (2) the effect of the stimulus context peculiar to the method of constants tends to obscure the pitch changes which take place.

Kinds of judgments.—In establishing the relationship between tonal duration and pitch qualitatively we wish to determine what frequency of a short duration tone has the same pitch as a longer duration tone with some stated pitch. In other words, we seek to find points of subjective equality. In the method of constants, we determine points of subjective equality inferentially, i.e., we ask the *S* to make judgments of differences (high or low) and then compute the points of subjective equality determining the frequency at which high and low judgments are equal in number. In the method of average error, however, the *S* makes equality judgments directly. The point of subjective equality then becomes the average of a number of such equality judgments.

In terms of the kind of judgment made, then, the method of average error is the more relevant to our problem and we feel, therefore, that the data from this method better represent the relation between pitch and duration.

Effect of stimulus context.—Data of Koester and Schoenfeld (4) suggest that in the method of constants the *S* tends to make his judgments in terms of a standard which is the mean of all the stimuli occurring during the run up to the time of making the judgment. We will point out in a moment how this would tend to obscure the amount of pitch change actually taking place as a function of duration. First, however, let us review the evidence offered by Koester and Schoenfeld to support such an hypothesis.

These authors were primarily interested in the effect of two or more interacting pitch stimuli on the C. E. function in pitch discrimination. Thus, in an experimental run they used several standards with which a number of comparison stimuli were paired. The *S*s were asked to judge whether the comparison tone was higher or lower than the standard with which it was paired (method of constants). An analysis of the data shows pronounced interaction effects. Thus, after stimulation at a high standard level (1100 cps) the judgments with respect to a lower standard level (1000 cps) were preponderantly low, and conversely, after stimulation at a low standard level (1000

cps) the judgments with respect to a higher standard (1100 cps) were preponderantly high. The effect was to force the points of subjective equality for the standards in the experimental run toward a mean of the series of stimuli in the run. In other words, the point of subjective equality for the lower standard tended to be higher and the point of subjective equality for the higher standard tended to be lower. The deduction was, then, that the *S* was making his judgments not with respect to the standard presented at the time but with respect to a standard which is determined by all of the interacting stimuli in the experimental run up to the time of making the judgment.

This information is useful in interpreting the data obtained by us with a method of constants, if we stop to consider the actual procedure used in a method of constants. In our experiment, the *S* heard a tone of 500 msec. duration followed, after a short silent interval, by a short tone. The frequencies of these short comparison tones were selected so that there were as many frequencies above that of the standard as below that of the standard. The mean of all these tones, then, was the frequency of the standard. If there is a constant error, however, these short tones are not symmetrically distributed around the *pitch* of the standard. Since the point of subjective equality tends toward the mean of all stimuli heard, it tends toward the frequency of the standard, not the pitch of the standard.

It might be said that the use of a method of constants, with all comparison tones symmetrically distributed about the frequency of the standard tone, tends to minimize constant errors. On the other hand, the *E* can, by proper selection of the comparison stimuli, determine the magnitude of the constant error. If, for example, all comparison stimuli have a higher frequency than the standard, then the point of subjective equality will be high. If all comparison stimuli are low, then the point of subjective equality will be low.

In a method of average error the *S* usually has complete control over the frequencies of the comparison stimuli. In our experiments he did. The *S* will at various times have the pitch of the comparison tone higher than that of the standard tone, and at other times it will be lower. In general, however, we can expect that the *S* will hear as many tones higher in pitch as lower in pitch with respect to the pitch of the standard tone. In other words, the *S* tends to distribute the tones symmetrically about the pitch of the standard, not its frequency. In fact, all judgments of the *S* must be in terms of the heard pitch, since he has no idea what the actual frequency is. The effect of context will still be present in a method of average error, but the context will be determined in terms of the actual pitch equalities of tones, which is what we are trying to determine. The *S*'s point of subjective equality will in no way be determined by the

arbitrary selection of values for the comparison stimulus which the *E* makes.

Whatever effects may exist, either in a method of constants or average error, due to the fact that all stimuli heard previously affect any single judgment, it is better if the *S* determines what stimuli he hears rather than the *E*. It is, after all, the psychology of the *S* we are concerned with, not the psychology of the *E*. Because of these relations, we again feel that the method of average error provides data which are more representative of the true relation between duration and pitch.

Variability

That the functions in Figs. 3 and 4 are not precise is indicated by the variability in judgments at each of the comparison stimulus durations, especially at the shorter durations. The envelope of shaded area about the line of best fit in Figs. 3 and 4 gives some indication of this variability in terms of the standard deviation of all judgments about the mean. This variability is probably due to a combination of the difficulty of judging pitch at very short durations and the inexperience of the *S*s.

A tone of only six msec. duration has very little pitch character to it. In the case of 1000 or 4000 cps of six msec. duration the tones are just beyond the click-pitch threshold, i.e., they are just beyond the point at which a predominant sensation of click has some pitch to it Since the pitch, in terms of which the *S* makes his judgment, is at a minimum at this duration, we would expect greatest variability at this point This expectation is borne out by the results.

As the duration of a tone is increased, the pitch becomes more specific. At 12 msec. a 1000 or 4000 cps tone is just beyond the tone-pitch threshold, i.e., beyond the point at which tone is just detectable between the on- and off-clicks. Because there is more pitch in a 12 msec. tone than in a six msec. tone, we would expect that judgments involving pitch would be more stable at the former duration than at the latter. Examination of the envelope of variability in Figs. 3 and 4 shows this to be the case.

It can be seen from these figures that with still further increases in duration, with consequent increase in the amount of pitch, variability decreases—remaining relatively constant beyond 100 msec. This increase in the variability of pitch matching with a decrease in duration is certainly in line with the previously reported measures of differential sensitivity for pitch as a function of duration.

Our data show one condition which presents a paradox. Previously we had shown that the click-pitch threshold for a tone of 250 cps was greater than six msecs. In other words, *S*s had stated that there was no pitch in a six msec. tone at a frequency of 250 cps. In this experiment, regardless of this fact, we asked *S*s to match the pitch of this tone to a tone of longer duration. That they were able to do so with some degree of consistency is indicated in the figures presented. Such a tone contains only one and a half complete cycles, and a pitch match would seem almost impossible.

That *S*s were, nevertheless, able to make a pitch match is the paradox, one which we are unable to explain adequately. The only reasonable explanation at all is that if the *S*s increased the frequency of the six msec. tone, it had a pitch which was distinctly different from that of a 250 cps tone. It is possible, then, that these judgments indicate the highest frequency for the short tone which did not produce a pitch clearly different from that of the 250 cps tone.

Theoretical Interpretations

The physical nature of short tones.—A very short tone is not a pure tone. Many frequencies, adding together in proper phase relations, are required to produce a short tone of any given period of oscillation and duration. The exact nature of the physical spectra of short tones was described in some detail in the previous paper (**3**), and for our purposes here, the following relations are all that are necessary.

Any short tone is composed of a band of frequencies distributed symmetrically about the frequency of oscillation, and most of the energy is contained in one main band whose frequency limits are defined by $f_0 \pm 1/d$, where f_0 is the frequency of oscillation, and d is the duration of the tone in milliseconds. Within this main band, the maximum energy occurs at the frequency of oscillation, and the rest of the energy is distributed equally on both sides of this central frequency, until no energy occurs at the limits of the main band. The energy in the other bands, farther removed from the central frequency, is very small compared to the energy in the main band and can, for most purposes, be ignored.

When short tones are used, then, the ear is not responding to a pure tone, but rather to a band of frequencies.

Response to the geometric mean.—Differential sensitivity to frequency is a relative matter; that is to say, the difference limen for frequency tends to be constant only if we measure the limen on a logarithmic scale of frequency. Likewise, it has been shown that equal units of subjective pitch are more closely related to equal logarithmic units of frequency than to equal arithmetic units of frequency. For these reasons, then, and the many related facts, we would expect the ear to respond to the geometric mean of a band of frequencies, rather than to the arithmetic mean, since the geometric mean of a ban of frequencies is the mean of those frequencies on a logarithmic scale. We expect this type of response on the assumption that the ear would average pitches rather than frequencies.

There are many ways of computing a geometric mean for bands of frequencies which are continuous and in which all frequencies do not have equal representation on the intensity scale. But regardless of the particular method chosen, the geometric mean is always less than the arithmetic mean. We would, then, expect that the pitch of short

tones would always be less than the pitch of a long tone. The shorter the tone, the wider the band of frequencies involved in that tone, and the lower the geometric mean with respect to the arithmetic mean. Thus, we would also expect that the shorter the tone, the lower the pitch of that tone would be as compared to some tone of longer duration.

It would be interesting to compute geometric means for many of the tones used in these experiments to see if the apparent pitch of the short tone approximates the pitch of the geometric mean of the band of frequencies involved. Such calculations would be, however, rather fruitless since so many arbitrary assumptions must be made in the calculations. For example, what band of frequencies shall we use in our computations? We do not know the effective bandwidth of energy as far as the ear is concerned, and the wider the band of energy we use to compute the geometric mean, the lower it will be. Likewise, should we use a geometric mean in which each frequency component has been weighted according to its amplitude, or simply concern ourselves with the total band of frequencies involved? These problems, for all practical purposes, make it impossible to determine precisely whether the ear is responding to the geometric mean of a band of frequencies. The *general* expectation that the pitch of a short tone should be lower than the pitch of a long tone is confirmed, however.

Differential sensitivity of the ear to frequencies.—Another factor which enters into the problem of the pitch of short tones is the fact that the ear is not equally sensitive to all frequencies in the audio spectrum. At any given intensity, frequencies in the middle range will appear to be louder than either higher or lower frequencies. Thus, we can not concern ourselves entirely with the physical spectrum of short tones: we must be concerned as well with the psychological spectrum, and with the concept of a psychological geometric mean. Under certain conditions, for example, some of the frequency components in a short tone may be completely inaudible, while others will be quite loud. Such differential sensitivity will certainly affect the apparent pitch of a tone. Once again we can not make precise predictions, but we can make a general sort of prediction which checks with our data within limits.

At low frequencies, the higher frequencies in any given band will be louder than the lower frequencies in that band. At high frequencies, on the other hand, the ear is more sensitive to the lower frequencies of any given band than to the higher frequencies. We would predict, then, that the pitch of short tones at the low frequencies would be relatively higher than the pitch of short tones at high frequencies.

As an illustration, we have computed the geometric means for tones lasting six msec. at 250 and 4000 cps. We have arbitrarily computed these as the geometric midpoint of a band of frequencies defined by $1/d$. In the case of 250 cps, this band was 167–333 cps. and in the case of 4000 cps, the band was 3917–4083. The geometric mean in the former case is 236 cps, and in the latter case 3999 cps. In percentage terms, the pitch of these tones should then be −5.6 percent for 250 cps, and −0.025 percent for 4000 cps. If, however, the sensitivity of the ear has its effect, then the pitch should be relatively higher for the lower frequency, and relatively lower for the higher frequency. Reference to Fig. 3 will show that this expectation is confirmed: at 250 cps, the pitch drop is only 4 percent, and is 2.2 percent at 4000 cps.

There are other deductions which can be made with this line of reasoning. One of these is that the pitch of short tones at lower frequencies should become relatively more positive at lower intensities, since equal loudness contours become steeper at lower intensities. This expectation is likewise confirmed. At 4000 cps, however, pitch tends to become positive at a lower intensity, which is exactly opposite to what would be expected. In general, however, it seems as though the ear is responding to some psychological geometric mean. A better way of phrasing this statement, perhaps, is to say that the ear seems to respond to the midpoint of the total range of the psychological stimulus.

Summary and Conclusions

The purpose of these experiments was to determine the direction and magnitude of pitch changes as a function of duration over a wide range of durations, frequencies, and intensities. Both a method of constants and a method of average error were used, and the major difference between the methods was in terms of the magnitude of the effect. This problem is discussed, and results obtained by the method of average error are accepted as being more valid. Further results and discussion indicate:

1. As the duration of tones is decreased, there is a tendency for all tones to lose pitch.

2. The amount of pitch loss is related to intensity and frequency:

a. Pitch loss is greatest for high intensities, and is somewhat less for lower intensities.

b. Low frequencies show the greatest amount of pitch loss.

c. High frequencies at low intensities show actual pitch gain at short durations.

3. Many of the relationships found can be explained in terms of the ear's responding to the geometric mean of all the frequency components involved in a short tone.

(Manuscript received for immediate publication May 14, 1948)

References

1. Békésy, G. v. Zur Theorie des Hörens. Über die eben merkbare Amplituden- und Frequenzänderung eines Tones. Die Theorie der Schwebungen. *Physik. Z.*, 1929, **30**, 721–745.
2. Bürck, W., Kotowski, P., & Lichte, H. Der Aufbau des Tonhöhenbewusstseins. *Elek. Nachr.-Techn.*, 1935, **12**, 326–333.
3. Doughty, J. M., & Garner, W. R. Pitch characteristics of short tones. I. Two kinds of pitch threshold. *J. exp. Psychol.*, 1947, **37**, 351–365.
4. Koester, T., & Schoenfeld, W. N. The effect of context upon judgments of pitch differences. *J. exp. Psychol.*, 1946, **36**, 417–430.
5. Postman, L. Time-error as a function of the method of experimentation. *Amer. J. Psychol.*, 1947, **60**, 101–108.
6. Stevens, S. S., & Davis, H. *Hearing: its psychology and physiology.* New York: John Wiley and Sons, Inc., 1938.
7. Turnbull, W. W. Pitch discrimination as a function of tonal duration. *J. exp. Psychol.*, 1944, **34**, 302–316.